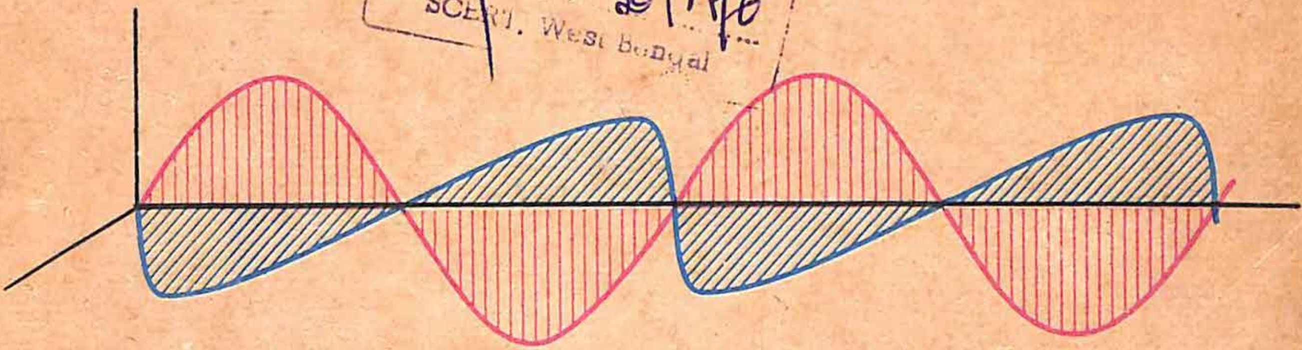
PHYSICS

CLASS XII
PART II



Kaṇāda ★ Archimedes ★ Kepler ★ Galileo ★ Newton
Henry Cavendish ★ Charles Coulomb ★ Thomas Young
André-M Ampère ★ C F Gauss ★ Amedeo Avogadro
Michael Faraday ★ James P Joule ★ Lord Kelvin
J C Maxwell ★ Ludwig Boltzmann ★ J J Thomson
Max Planck ★ Marie Curie ★ Robert A Millikan
E Rutherford ★ Albert Einstein ★ Niels Bohr ★ de Broglie
Erwin Schrödinger ★ Werner Heisenberg ★ Paul Dirac
C V Raman ★ M N Saha ★ S N Bose ★ W Pauli
Enrico Fermi ★ Lev D Landau ★ S Chandrasekhar
Edwin Hubble ★ John Bardeen ★ Richard Feynman

PHYSICS

Volume II Part II

A Textbook for Class XII

AUTHOR TEAM

CHAIRMAN

T. V. RAMAKRISHNAN
Indian Institute of Science, Bangalore

MEMBERS

ARVIND KUMAR
Homi Bhabha Centre for Science Education, Bombay
A. N. MAHESHWARI
National Council of Educational Research & Training, New Delhi
N. MUKUNDA
Indian Institute of Science, Bangalore
RAJARAM NITYANANDA
Raman Research Institute, Bangalore
J. RAMAKRISHNA
Indian Institute of Science, Bangalore
S. RAI CHOUDHURY
Delhi University, Delhi
D. P. SENGUPTA
Indian Institute of Science, Bangalore
S. V. SUBRAMANYAM
Indian Institute of Science, Bangalore
C. V. VISHVESHWARA
Indian Institute of Astrophysics, Bangalore

PHYSICS CURRICULUM GROUP

Department of Education in Science and Mathematics, NCERT

R. N. MATHUR
B. K. SHARMA
R. JOSHI
V. P. SRIVASTAVA

PHYSICS

Volume II Part II

A Textbook for Class XII

ARVIND KUMAR
A. N. MAHESHWARI
N. MUKUNDA
RAJARAM NITYANANDA
J. RAMAKRISHNA
T. V. RAMAKRISHNAN
S. RAI CHOUDHURY
D. P. SENGUPTA
S. V. SUBRAMANYAM
C. V. VISHVESHWARA

राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING

First Edition
March 1989 Chaitra 1911

Fourth Reprint Edition
October 1993 Kartika 1915

Revised Edition
October 1995 Kartika 1917

PD 100T GR

Cover : Designed by Professor D. P. Sengupta and Professor Indraneel Sen, Indian Institute of Science, Bangalore. Executed by Mr. A. S. Prabhakar, USE COMP, Bangalore.

OFFICES OF THE PUBLICATION DEPARTMENT, NCERT

NCERT Campus	CWC Campus	Navjivan Trust Building	CWC Campus
Sri Aurobindo Marg	Chitlapakkam, Chromepet	P.O. Navjivan	32, B.T. Road, Sukchar
NEW DELHI 110016	**MADRAS 600064**	**AHMEDABAD 380014**	**24 PARGANAS 743179**

Rs 72.50

Published at the Publication Division by the Secretary, National Council of Educational Research and Training, Sri Aurobindo Marg, New Delhi 110 016, lasertypeset at the Indian Institute of Science, Bangalore 560 012 and printed at J. K. Offset Printers, 315, Gali Garahia, Jama Masjid, Delhi 110 006

CONTENTS

CHAPTER 7

Electromagnetic Induction

CHAPTER 8

Voltages and Currents Varying with Time

CHAPTER 9

Electromagnetic waves

CHAPTER 10

Wave Optics

CHAPTER 11

Ray Optics and Optical Instruments

CHAPTER 12

Electrons and Photons

CHAPTER 13

Atoms, Nuclei and Molecules

SCIENCE RELATED VALUES

Curiosity, quest for knowledge, objectivity, honesty and truthfulness, courage to question, systematic reasoning, acceptance after proof/verification, open-mindedness, search for perfection and team spirit are some of the basic values related to science. The processes of science, which help in searching the truth about nature and its phenomena are characterised by these values. Science aims at explaining things and events. Therefore to learn and practise science :

* **Be inquisitive about things and events around you.**
* **Have the courage to question beliefs and practices.**
* **Ask 'what', 'how' and 'why' and find your answers by critically observing, experimenting, consulting, discussing and reasoning.**
* **Record honestly your observations and experimental results in your laboratory or outside it.**
* **Repeat experiments carefully and systematically if required, but do not manipulate your results under any circumstance.**
* **Be guided by facts, reasons and logic. Do not be biased in one way or the other.**
* **Aspire to make new discoveries and inventions by sustained and dedicated work.**

CHAPTER **7**

Electromagnetic Induction

7.1 Introduction

Electricity and magnetism were considered separate, unrelated phenomena for a long time. You have studied the former in Chapters 1 and 2, where electric charges at rest are considered. Chapter 6 similarly describes magnetism as being due to permanent magnets which were known to mankind for millenia. In Chapter 3, electric charges in motion, or electric currents were discussed. In the early decades of the nineteenth century, experiments by Oersted, Ampere and others using electric currents established the fact that electricity and magnetism are interrelated; they found that *moving electric charges* or electric currents cause magnetic fields which can, for example, deflect a magnetic compass needle (Chapter 5). This naturally raises the question: Is the converse effect possible? Namely, can a moving magnetic field give rise to electricity? This ques-

tion occupied the attention of Michael Faraday, the great British scientist, for many years. After a variety of experiments, he found in 1831 that a moving magnetic field can indeed give rise to an emf. Independently, the effect was discovered by Joseph Henry in the USA at about the same time.

7.2 Faraday's experiments

We describe now some experiments of the sort that Faraday and Henry performed.

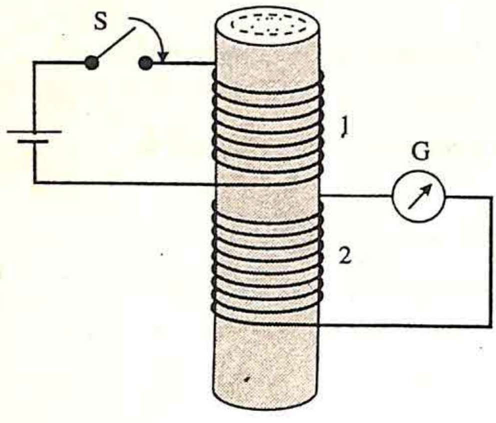


Figure 7.1: If the switch S in coil 1 is closed or opened, the galvanometer G shows a deflection which persists only for a very short time.

Experiment 1

Two coils of insulated copper wire were wound on a wooden rod as shown in Fig. 7.1. The coils are not connected to each other. In one coil (coil 1), there is a switch and a cell. The second coil (coil 2) has a galvanometer in the circuit.

Faraday observed that the needle of the galvanometer connected to coil 2 was deflected as soon as the switch in coil 1 was

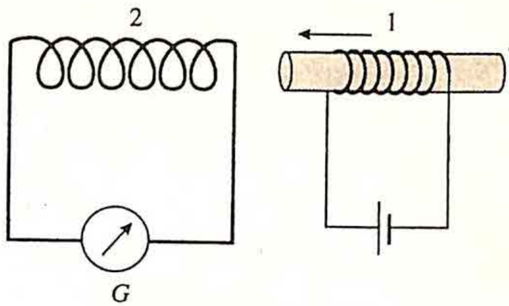


Figure 7.2: If coil 1 carrying a direct current is brought towards coil 2, or moved away from it, the galvanometer G in coil 2 shows a momentary deflection.

closed. The same thing happened when the switch was opened but this time the galvanometer needle was deflected in the opposite direction. When the switch was kept closed and a steady current flowed in coil 1, the galvanometer did not show any deflection. The deflections were momentarily seen only when the switch was opened or closed. When Faraday replaced the wooden rod by an iron rod, the galvanometer showed larger deflections on closing and opening of the switch.

Experiment 2

In this experiment, coil 1 carried a steady current, and as coil 1 was being brought close to coil 2, the galvanometer showed a deflection (Fig. 7.2). As coil 1 was moved away, the galvanometer showed a deflection again, but this time in the opposite direction. The deflection lasted as long as coil 1 was in motion. If instead of moving a coil carrying a current, a bar magnet was moved towards coil 2 or away from it (Fig. 7.3), the same effects were observed. In both the experiments (Figs. 7.1 and 7.2), larger deflections of the galvanometer were observed if the coil 1 carried larger currents. In ex-

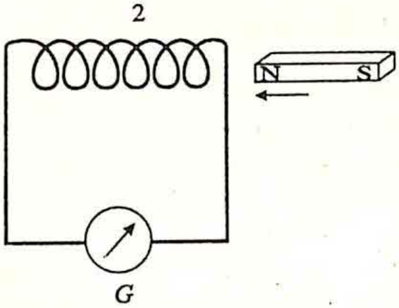


Figure 7.3: If a magnet is suddenly brought close to coil 2 or moved away, the galvanometer shows a deflection as in Fig. 7.2.

periment 2 (Figs. 7.2 and 7.3) the deflection was larger if the coil or the bar magnet was moved faster. The coil 1 in Fig. 7.2 carries a steady electric current and produces a magnetic field so long as a current flows. Such an arrangement is generally known as an electromagnet.

7.3 Faraday's law

From these observations Faraday came to conclude that an emf was being *induced* in coil 2, when the *magnetic flux linked* by it was being *changed* with time either by

(i) switching on or switching off a current in coil 1 (Experiment 1, Fig. 7.1) or

(ii) moving an electromagnet (Fig. 7.2) or a magnet (Fig. 7.3) close to or away from coil 2. (Experiment 2).

and it was this induced emf which caused a current to flow in coil 2, through the galvanometer. In order to connect the emf induced and the magnetic flux, we recall the definition of magnetic flux in Chapter 6 (Section 6.2). Suppose a uniform magnetic field **B** passes through a plane of area A at an angle θ with the normal to the plane, the magnetic flux is

$$\phi = BA\cos\theta \tag{7.1}$$

This relation can be easily extended to curved surfaces and nonuniform fields.

In the experimental arrangement described in Figs 7.1 to 7.3, the coil 2 has many turns N and let us assume that the same flux ϕ passes through each turn. Thus the total flux linked by N turns of the coil is $N\phi$. Faraday concluded from his experiments that the emf induced in the coil 2 is $\mathcal{E}$ where

$$|\mathcal{E}| = \frac{\mathrm{d}}{\mathrm{d}t}(N\phi). \tag{7.2}$$

Equation 7.2 gives only the magnitude of the induced emf $\mathcal{E}$. It does not indicate the direction of the induced emf. We turn to this aspect now.

7.4 Direction of the induced emf: Lenz's law and conservation of energy

In Fig. 7.3 of experiment 2, if instead of bringing the North pole close to coil 2, we bring the South pole close to it, the deflection of the galvanometer would be in the opposite direction. When the N-pole is brought *towards* the coil, the current induced flows in the counterclockwise direction as seen by the advancing North pole of the bar magnet. Using the corkscrew rule, you can see that this current produces a magnetic field in the direction of the forefinger which opposes the magnetic field of the bar magnet (see Fig. 7.4). The coil side facing the advancing North pole also becomes a North pole and like poles repel each other. If the South pole of the bar magnet is brought towards the coil the current flows in clockwise direction and a South pole is induced on the coil side facing the advancing South pole of the bar magnet, once again repelling the advancing South pole. (The inset in Fig. 7.4 helps one to remember the direction of the current and the induced polarity).

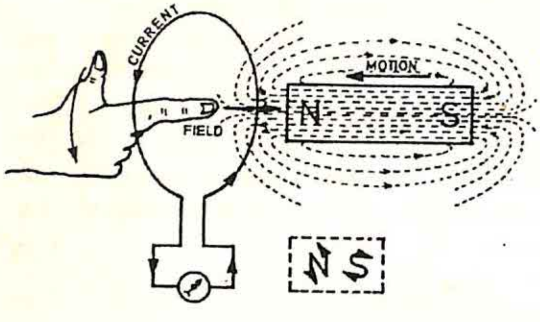


Figure 7.4: A N-pole of a bar magnet pushed towards a coil induces a N-pole on the face of the coil and experiences repulsion.

One can conclude from these observations that the emf induced in the coil causes a current to flow which opposes the change (here the motion of the magnet) that causes the emf. This is a very general feature of such induced emf's that was recognized by Lenz in 1835. The general result called Lenz's law states that "*the induced current due to the induced emf always flows in such a direction as to oppose the change causing it*".

Faraday's law, Eq. (7.2) can now be combined with Lenz's law to give what is commonly known as Faraday-Lenz's law (often called Faraday's law). This relates the induced emf in a coil with the rate of change of magnetic flux linkage and states that "**the emf induced in a coil is equal to the negative of the rate of change of magnetic flux linked with it**", or

$$\mathcal{E} = -\frac{\mathrm{d}\phi_{total}}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}(N\phi) = -N\frac{\mathrm{d}\phi}{\mathrm{d}t} \quad (7.3)$$

Here ϕ_{total} is the total magnetic flux linked by the coil of N turns. When a coil is closely wound, the same magnetic flux may be assumed to pass through each turn and $\phi_{total} = N\phi$ (since N or the number of turns does not change with time, Eq. (7.3) is valid).

We now argue that Lenz's law is just an expression of the general law of conservation of energy. Let us assume that moving the *North pole* of a magnet towards a closed coil (Fig. 7.4) induces a current in such a direction that the end of the coil (facing the approaching North pole) becomes a *South pole* and not a North pole. The magnet will then be drawn towards the coil instead of being repelled. In other words, give a little push to the magnet and the magnet continues to accelerate towards the coil on its own! Now the current I induced in the coil 2 having a resistance R will cause I^2R loss or heat loss in the coil. The energy required to move the magnet and the heat energy loss in the coil seem to come from nowhere! Surely this cannot be true since it violates the principle of energy conservation. If, on the other hand, the induced current produces a North pole and opposes the motion, then work has to be done to push the magnet against the coil. It is this mechanical work which causes a current to flow in the coil against its resistance R and supplies the energy for the heat loss. *The mechanical work done is converted to electrical energy which produces the heat energy.* So, everything seems to be right if we accept Lenz's law that *the induced current flows in a direction so as to oppose the change causing it.* Otherwise, the principle of energy conservation is violated.

7.5 Discussion of Faraday's law

7.5.1 Units and size of induced emf

We now discuss Faraday's law of electromagnetic induction using several examples. Before that, we briefly go into the question of the size of electromagnetic induction and of units. Equation (7.3) is correct as it stands

in the SI units, i.e. it gives the induced emf in volts if the magnetic flux is in tesla × square metres and time is in seconds. The former is the natural unit for magnetic flux which is dimensionally, magnetic field times area. The magnetic flux is such a commonly occurring quantity that its SI unit tesla metre^2 (or Tm^2) is given a separate name, called weber and abbreviated as Wb. We thus say that if the magnetic flux linking a *single turn* changes at the rate of one weber per second, the induced emf is one volt.

Let us try to get some idea of the size of electromagnetic induction. A flux of one weber is rather large. It corresponds to a field of one tesla (which is quite large) going through a coil of one turn having an area 1 m^2. If this flux changes in one second from 1 weber to zero, an emf of one volt is induced. One volt is a low voltage. The means to produce larger voltages becomes obvious if we examine Faraday-Lenz's law closely. Since

$$\mathcal{E} = -N\frac{d\phi}{dt}$$

where ϕ is the magnetic flux in webers enclosed by a single turn coil. $\mathcal{E}$ will increase if we

(i) increase N, or the number of turns of the coil

(ii) increase the rate of change of ϕ.

If instead of one turn, the coil has N=1000 turns (area is still $1m^2$) and the enclosed field of 1T decreases to zero in (1/50) of a second, then

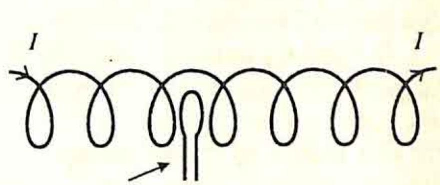
$$|\mathcal{E}| = N\frac{d}{dt}(BA)$$

$$|\mathcal{E}| = 1000 \times \left(\frac{1}{1/50}\right) \times 1 \times 1$$

$$= 50,000\text{V}$$

which is fairly large. If the field is reduced to 0.1 T and the area of the coil is only 0.2 m^2, 1000 V will be produced. This is nearly five times the voltage at which we use electric power at home. In the discussion of Faraday's law of electromagnetic induction, we always say that an emf is induced, not a voltage or a current. This is indeed true; for example when in the experimental setup described in Fig. 7.2, the coil 2 has a break, no induced electric current passes through it, but a potential difference does develop between its broken ends as coil 1 moves towards it.

Example 7.1: A long solenoid of 10 turns/cm has a small loop of area 1 sq. cm. placed inside with the normal of the loop parallel to the axis. Calculate the voltage across the small loop if the current in the solenoid is changed at a steady rate from 1A to 2A in 0.1s, during the duration of the change.



Answer: The magnetic field inside the sonenoid changes at the rate:

$$\begin{aligned}\frac{dB}{dt} &= \left(\mu_0 n\frac{dI}{dt}\right)\\ &= 4\pi \times 10^{-7} \times (10 \times 100)\times \left(\frac{1}{0.1}\right) \text{Wb m}^{-2}\text{s}^{-1}\\ &= 12.57 \times 10^{-3}\text{Wb m}^{-2}\text{s}^{-1}\end{aligned}$$

The induced voltage which is equal to the

rate of change of flux, therefore is

$$\begin{aligned} V &= \left(\frac{dB}{dt}\right) \times (\text{Area}) \\ &= (12.57 \times 10^{-3}) \times 10^{-4}\,\text{V} \\ &= 12.57 \times 10^{-7}\,\text{V} \end{aligned}$$

7.5.2 Motional emf

Fig. 7.5 represents an electromagnet. The coil C_1 connected to a battery carries a current in the direction shown by the arrows. Using the corkscrew rule we can check the direction of the magnetic field and label the poles as N (the pole from which the magnetic field lines emerge) and S (the pole into which the magnetic field lines enter). Let a coil C_2 with a galvanometer in the circuit be placed in the gap between the N and S poles. This is shown in an enlarged version in Fig. 7.6. The crosses show the direction of the magnetic field emerging from the North pole and entering the South pole *into* the paper. If we turn off the switch, current stops flowing in the coil C_1 and the magnetic field ceases to exist. The galvanometer in the coil C_2 shows a momentary deflection since the magnetic flux linked by it has changed.

If instead of opening the switch, we allow a steady current in C_1 to flow so that the magnetic field in the gaps is as before, but move the coil *inside* the poles, do you expect a voltage to be induced in coil C_2 according to Faraday's Law?

Will the galvanometer be deflected? The answer is 'No'.

So long as the coil is always fully inside the pole (see Fig. 7.6) the flux linked with the coil remains unaltered although you may shift it around. Since the flux linkage does not change, no emf is induced.

If, however, the coil is partly inside the poles and partly outside as in Fig. 7.7, and

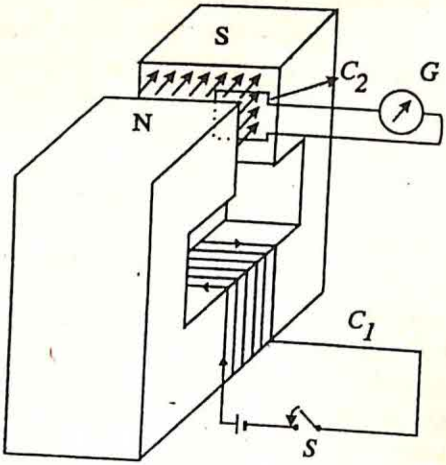


Figure 7.5: An electromagnet, with a plane coil connected to a galvanometer between its pole pieces.

you pull the coil C_2 away from the magnetic field, a current *will* flow momentarily as will be shown by the deflection of the galvanometer needle. In this case you can clearly see that the flux enclosed by the coil C_2 decreases (i.e. changes) as the coil is pulled out, and therefore an emf is induced. We see from these examples that a change in the linked magnetic flux which induces an emf may be produced in two ways:

(i) There is no physical motion of either the source of magnetic field or of the coil through which the magnetic field passes but the magnetic field itself changes with time. This may be caused by changing the current that produces the magnetic field (Recall Faraday's experiment 1 described in Fig. 7.1).

(ii) The change is produced by relative motion of the source of the magnetic field and the coil through which the magnetic field passes. In Fig. 7.2 and 7.3 it is the magnetic field that changes with

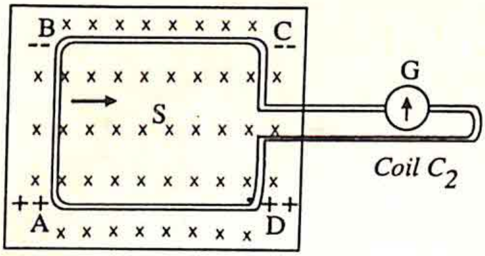


Figure 7.6: If a coil is moved entirely within a uniform magnetic field, it does not experience any change and no current is produced and the galvanometer shows no deflection.

time and in Fig. 7.7 the conductor moves through a fixed magnetic field, but in both cases the linked magnetic flux changes with time.

It is found experimentally that in either of these cases the induced emf is given by the same law i.e. Eq. (7.3), being equal to the time rate of change of linked magnetic flux. The voltage in the latter case is often called *motional emf.* We show now that the phenomenon of motional emf can be easily understood in terms of the magnetic Lorentz force on moving charges.

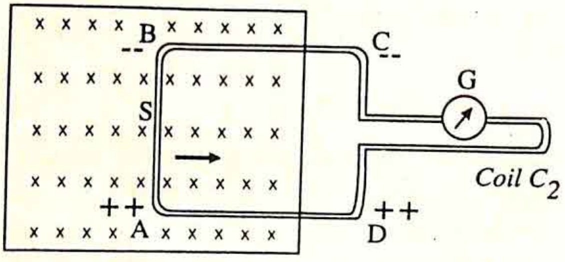


Figure 7.7: If a part of the coil is inside the field and a part outside and the coil is pulled out, the net flux linkage with time changes and a current flows. The galvanometer also shows deflection as the coil is moved.

7.6 Electromagnetic induction and Lorentz force

You have studied in Chapter 5, Section 6 that a charge q moving with a velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$ experiences a magnetic Lorentz force.

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) \tag{7.4}$$

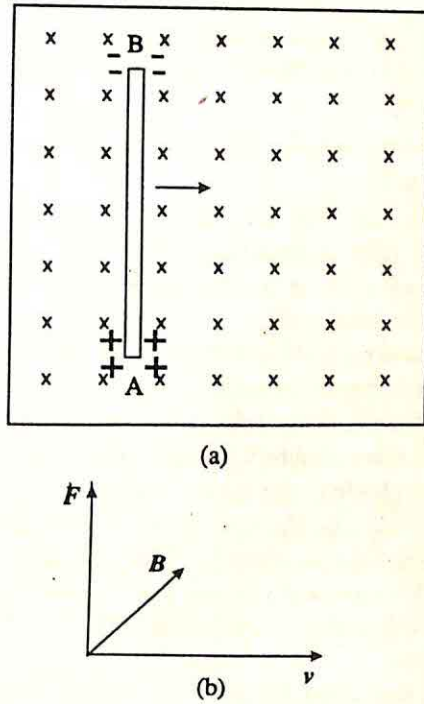


Figure 7.8: If a metal rod AB is moved inside a magnetic field, the Lorentz force causes electrons to move to the end B if the rod is moved as shown.

When we drag a metal, free electrons present in it gain velocity in the direction of motion. As soon as the metal enters a magnetic field, its free electrons experience Lorentz force perpendicular to both the magnetic field $\mathbf{B}$ and their velocity $\mathbf{v}$. The electrons under this force accumulate at one end, providing it negative polarity. The other end, deprived of electrons, becomes

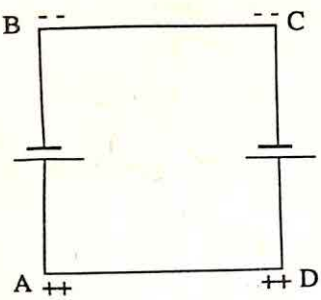


Figure 7.9: If two cells equivalent to the emf in AB and CD respectively, are connected as shown, no current flows.

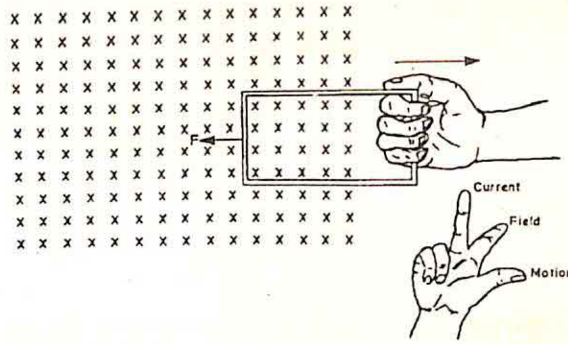


Figure 7.10: As the coil is pulled away from the field, motional emf causes a current to flow. The direction of the magnetic field, direction of motion and the direction of current can be determined using Fleming's right hand rule

positively charged. Fig. 7.8 shows this for a metal rod.

As the charges accumulate at the ends an electric field is produced. This field in turn produces a force on electrons in the metal which is easily seen to be opposite to the Lorentz force. When enough charges accumulate, this force cancels the Lorentz force and the free electrons do not drift anymore. In the case where the coil ABCD is completely inside the magnetic field and is pulled (Fig. 7.6) the charges accumulate both on AB and CD as shown. The coil ABCD is then like two cells of same polarity connected as shown in Fig. 7.9. In this case no current will flow.

On the other hand in the case of Fig. 7.7 where a section of the coil is outside the field, the emf induced in AB is not balanced by that in CD (as it is in Fig. 7.6), there is a net emf, and a current flows in the circuit.

7.7 Motional emf: a quantitative study

Let us look again at the rectangular single turn conductor ABCD being pulled through a magnetic field as shown in Figs 7.6 and 7.10. In the part AB, the magnetic Lorentz force on the free electrons in the conductor is evB in magnitude and points along AB. Thus the electrons preferentially accumulate at B. This continues as discussed above, till the electric field **E** due to this charge accumulation (electron excess or negative charge at B and deficit or positive charge at A) is such that the electrostatic force due to this exactly balances the magnetic Lorentz force. This clearly will happen for

$$E = vB. \tag{7.5}$$

This electric field exists along AB, so the electric potential difference between A and B is

$$V_{AB} = vB\ell = B\ell v. \tag{7.6}$$

This is the emf induced in the rectangular conductor ABCD due to its motion in the magnetic field.

Let us compare it with what one expects from Faraday's law. Suppose at a given instant a length x is inside the magnetic field (see Fig. 7.10). The magnetic flux linked with the rectangular conductor (of single turn) is

$$\phi = B\ell x. \tag{7.7}$$

Since x changes with time at a steady rate v, the rate of change of linked flux is

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = -B\ell\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right) = -B\ell v. \qquad (7.8)$$

The induced emf, according to Faraday's law of electromagnetic induction, is

$$\mathcal{E} = -\frac{\mathrm{d}\phi}{\mathrm{d}t} = B\ell v. \qquad (7.9)$$

This is indeed the emf as calculated from the principle of Lorentz force, Eq. (7.6). Note that ABCD is a coil with a single turn and hence N of Eq. (7.3) is unity.

We have thus found that the magnetic Lorentz force on a conductor moving in a magnetic field gives rise to the emf that was discovered by Faraday. If only the relative motion of the conductor and the magnetic field matters, then a magnetic field moving through a conductor with a velocity $\mathbf{v}$ should be equivalent in its effects to the reverse, namely the (same) conductor moving through the (same) magnetic field with a velocity $-\mathbf{v}$. It should thus produce the same emf. Further, if all that matters is the change of magnetic flux with time, this can be produced by a physically stationary source of magnetic field, but by changing with time the electric current causing the magnetic field. We cannot demonstrate these equivalences here further than we have done, but there is only one cause of induced emf. Roughly, one can say that a *time varying magnetic field produces an electric field.* The Faraday's law of electromagnetic induction, and the Lorentz force law on moving electric charge are two consequences of this basic effect.

To complete the discussion of the moving conductor in a magnetic field, we calculate the power loss P and the force F connected with this motion. Suppose the resistance of the square conductor loop is R. The current that flows is

$$I = (B\ell v/R) \qquad (7.10)$$

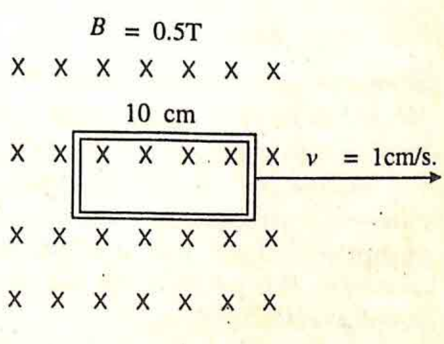
As the coil is pulled in the direction shown, an equal and opposite force, (shown as $\mathbf{F}$ in Fig. 7.10) resists the motion (Lenz's Law and also Newton's Law). The work done on the coil appears as heat due to the I^2R loss in the coil. If we assume that there is no mechanical frictional force present, all the power P spent in pulling the coil against $\mathbf{F}$ is expended as joule heat loss:

$$\begin{aligned} P &= I^2R = \frac{B^2\ell^2v^2}{R^2} \times R \\ &= \frac{B^2\ell^2v^2}{R} \end{aligned} \qquad (7.11)$$

Since power = force × velocity, the force $\mathbf{F}$ has a magnitude

$$F = \frac{P}{v} = \frac{B^2\ell^2 v}{R}. \qquad (7.12)$$

Example 7.2: A rectangular loop of sides 10 cm and 3 cm with a small cut is moving out of a region of uniform magnetic field of magnitude 0.5 T directed normal to the loop. What is the voltage developed across the cut if the velocity of the loop is 1 cm s^{-1} in a direction



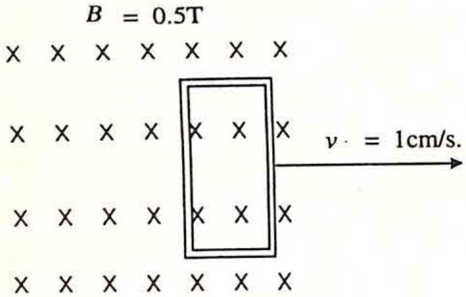
(i) normal to the shorter side

(ii) normal to the longer side of the loop?

For how long does the induced voltage last in each case?

(iii) Is a force required to pull the loop if it has a cut?

(iv) Find the force needed to pull the loop if it has no cut, and a resistance of 1 m Ω.



Answer: The emf induced in the gap is

$\mathcal{E}_1 = B\ell_1 v$.

(i) Here $\ell_1 = 3$ cm $= 0.03$ m.

$$v = 1\text{cm/s} = 0.01 \text{ m/s}.$$

The voltage induced is, therefore

$$\begin{aligned}\mathcal{E}_1 &= 0.5 \times 0.03 \times 0.01 \\ &= 0.15\,\text{mV}.\end{aligned}$$

Since the velocity is 1 cm/s, it takes 10 cm/(1cm/s) i.e. 10 seconds for the long side to come out of the field. So the voltage lasts for 10 seconds. [We have obviously assumed that the field abruptly changes from 0.5 T to zero, in solving this problem. In reality this is not exactly correct].

(ii) Here, we have

$$\begin{aligned}\ell_2 &= 10\text{cm} = 0.1\text{m} \\ \mathcal{E}_2 &= B\ell_2 v = 0.5 \times 0.1 \times 0.01\text{V} \\ &= 0.5\text{mV}.\end{aligned}$$

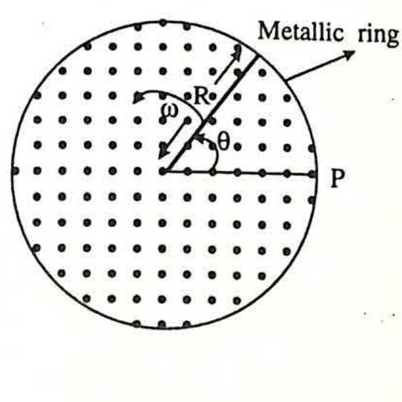
This voltage lasts 3 cm/(1 cm/s) = 3 s; this being the time the shorter or the 3 cm side takes to come out of the field.

(iii) Because of the gap, no current can flow. If there is no current, there cannot be any I^2R loss or heat produced. If we neglect friction, no force is required to pull the coil.

(iv) The force required in the case (Fig.(a)) is

$$\begin{aligned}F_1 &= \frac{B^2\ell_1^2 v}{R} = \frac{0.5^2 \times (0.03)^2 \times 0.01}{1 \times 10^{-3}} \text{ N} \\ &= 2.25 \times 10^{-3}\text{N}.\end{aligned}$$

Example 7.3: A conducting rod of 1 m length moves with a frequency of 50 rev/s, with one end at the centre and the other end at the circumference of a circular metallic ring of radius 1 m, about an axis passing through the centre of the coil perpendicular to the plane of the coil. A constant magnetic field parallel to the axis is present everywhere. What is the emf developed between the centre and the metallic ring? (B = 1.0 Wb/m^2).



Answer: There are no steady currents in this example. Separation of charges takes place. To calculate the emf, we can imagine a closed loop by connecting the centre with any point on the circumference, say P, with a resistor. The potential difference across the resistor is then equal to the induced emf and equals

$B \times$ (rate of change of area of loop).

If θ is the angle between the rod and the radius of the circle at P at time t, the area of the arc formed by the rod and the radius at P is:

$$\pi R^2 \times \frac{\theta(t)}{2\pi} = \frac{1}{2} R^2 \theta(t)$$

where R is the magnitude of the radius of the circle. Hence the induced emf is

$$B \times \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}R^2\theta(t)\right]$$
$$= \frac{1}{2} B R^2 \frac{\mathrm{d}\theta(t)}{\mathrm{d}t}$$
$$= \frac{1}{2} \times 1.0 \times (1^2) \times 50 \times 2\pi$$
$$= 157\mathrm{V}$$

[Note: $\frac{\mathrm{d}\theta}{\mathrm{d}t} = \omega = 2\pi\nu$]

7.8 Principles of ac generation; ac generators

We now describe briefly one extremely important application of electromagnetic induction. This is the generation of alternating currents or ac. Present day ac generation is a very highly evolved field, owing much of its initial impetus to Nikola Tesla, a Yugoslav inventor. We outline some basic principles here.

Consider, as in Fig. 7.11a, a rectangular loop rotating at a steady angular velocity ω about an axis AB passing through the plane of the loop. The loop is in a uniform field **B** as shown. In figures 7.11a, b, and c, we show the loop at different instants of time.

In Fig. 7.11a, the loop is normal to the field and cuts the largest flux ϕ where

$$\phi = B\ell b = BA,$$

In Fig. 7.11b, the loop moves through an angle θ and makes an angle $\{(\pi/2) - \theta\}$ with the flux lines. The area of the coil that *faces* the field is $\ell.b \ \cos\theta = A\cos\theta$.

In Fig. 7.11c, $\theta = 90°$ or the coil is parallel to the field and the area facing the field is $A\cos\theta = 0$.

If the coil is rotated with an angular velocity of ω, then $\theta = \omega t$, counting time t from the coil position in Fig. 7.11a. The flux cut by the coil at any position is

$$\phi = BA\cos\omega t$$

Using Faraday - Lenz's law

$$\mathcal{E} = -N\frac{\mathrm{d}\phi}{\mathrm{d}t} = -N\frac{\mathrm{d}}{\mathrm{d}t}(BA\cos\ \omega t)$$
$$= NBA\omega\ \sin\omega t \qquad (7.13)$$

where N is the number of turns of the coil. So if we rotate a coil of N turns in a magnetic field B, the voltage across the gap will not be a constant, but *will vary sinusoidally*, alternately becoming positive and negative (Fig. 7.12). The maximum value of the voltage is

$$V_m = BAN\omega.$$

If the frequency of rotation is ν revolutions per second

$$\omega = 2\pi\nu$$
$$V_m = 2\pi\nu BAN \qquad (7.14)$$

If we increase the frequency of rotation the magnitude of the generated voltage (V_m) and also the frequency of polarity change increases. This is alternating or ac voltage. If we increase the number of turns, the area

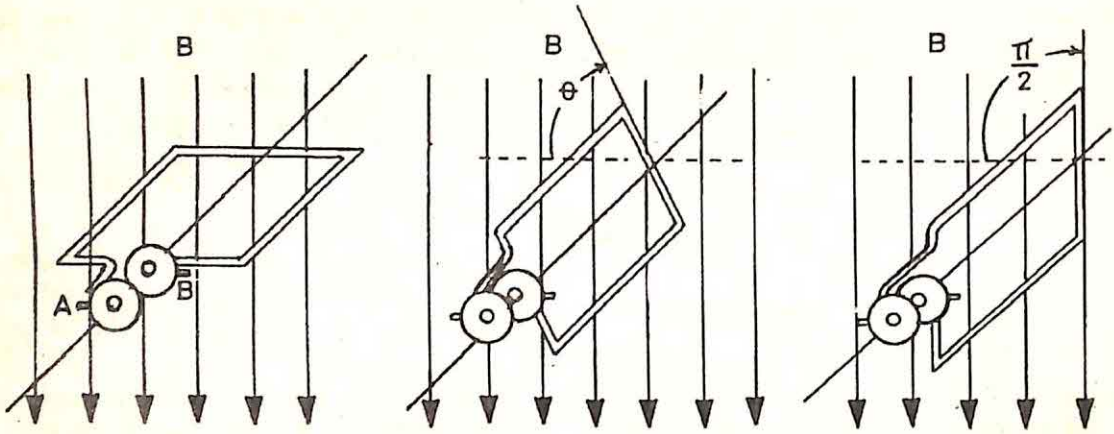


Figure 7.11: Principles of an ac generator. When a coil is rotated in the presence of a magnetic field, motional emf is produced across the brushes AB.

of the coil and the magnetic field B, the voltage generated increases. This is illustrated in Example 7.4.

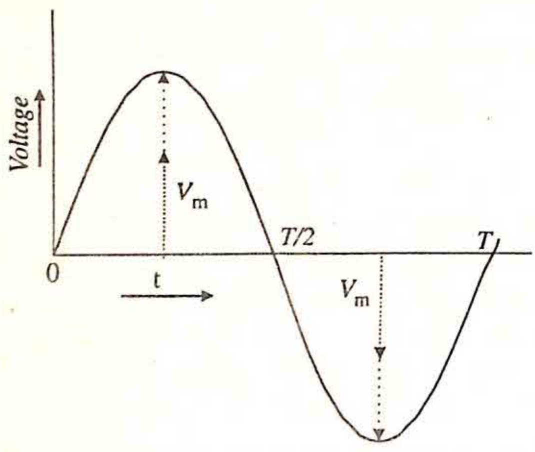


Figure 7.12: This represents the voltage generated across the brushes AB of Fig. 7.11. If the brush has positive polarity during the first half cycle $0 < t < (T/2)$ it has negative polarity during the next half cycle. If this polarity alternates 50 times in a second, the ac voltage is said to have a frequency of 50 Hz.

Example 7.4: A rectangular coil of length 1 m and width 0.5 m, and 10 turns is rotated at 50 revolutions per second (or at $\nu = 50$ Hz). The magnetic field within which the coil is rotated $B = 0.5$T. Calculate the peak value of the voltage generated across the ends of the coil.

Answer: The peak value of the voltage generated is

$$\begin{aligned} V_{max} &= 2\pi\nu BAN \\ &= 2\pi \times 50 \times 0.5 \times 1 \times 0.5 \times 10\text{V} \\ &= 785\text{V} \end{aligned}$$

So we see how we can generate high voltage starting from the simple laws of Michael Faraday. In fact, ac generators producing very large power are made these days. One generator producing 500 million watts (or megawatts, MW) of power at about 15 kV is common. Very large amounts of mechanical power are, therefore, necessary to rotate the rotor which produces the magnetic field inside enormously large coils. The rotors are rotated by what are called turbines. These turbines are driven either by water coming from a height in big pipes or by steam at high pressure obtained by heating water. The energy from the falling water is converted into electrical energy; such generators are called

hydroelectric generators. In thermal generators, the energy is obtained from burning coal (or from nuclear sources) and a single generator can provide power for 5 million 100 watt bulbs! In Fig. 7.11 we have seen how ac voltage is produced when a coil is rotated in a magnetic field. In real generators involving millions of watts of power, the coils are held stationary, and electromagnets are rotated.

Hydrogenerators run at slow speeds (about 600 rpm), and to produce an alternating voltage at say, 50 Hz, the rotor carries a large number of electromagnets.[1] A stream turbine rotor has an angular speed of 3000 rpm, so that with one electromagnet rotated at this speed, an ac voltage of $(3000/60) = 50$ rps or 50 Hz is produced.

7.9 Transformer

For many purposes, it is necessary to change electricity at one voltage to another. Domestic electric power in our country as well as most countries is used at about 220 V to 240 V at 50 Hz. In the U.S and Canada ac power is supplied at 110 V at 60 Hz.

It is not economical to generate large amounts of electricity at such low voltages, leave alone transport them over long distances. As water is supplied in huge pipes from a reservoir at a pressure to reach all parts of a city, electricity is also to be supplied from the source at high pressure or voltage and we should need conductors of enormous size to transport bulk power. If a large current flows through a conductor of small cross-section (i.e. large resistance) joule heating produced due to I^2R loss is so large that the conductor would melt. Besides, conductors of small sections and therefore high resistance, cause the voltage to drop and the flow of current is impeded. But how large can we make the conductors in order to make their resistance sufficiently low for delivering large amounts of current? Let us examine this.

Electricity is conveniently generated at 11 kV to 15 kV and generating 210 MW from a single generator is common. 210 MW of power at say, 15 kV involves nearly 9000 amperes of current (unlike in dc, ac power is not always given as $P = VI$ as we shall see in Chapter 8. It is nevertheless *proportional* to VI). We would need copper conductor of about 8 cm diameter to conduct this current without overheating. The weight of this conductor per kilometre would approximately be 32,000 kgs and its cost would be enormous. If on the other hand we could transmit the same power at a higher voltage the current would obviously be less. In fact 210 MW transmitted at 220 kV would involve approximately 600 amperes which is more manageable.

In order to have the same I^2R loss for transporting a certain amount of power P over a certain distance ℓ, $I_1^2R_1 = I_2^2R_2$ or $I_1^2\ell\rho/A_1 = I_2^2\ell\rho/A_2$ and therefore,

$$\frac{A_1}{A_2} = \frac{I_1^2}{I_2^2} = \frac{V_2^2}{V_1^2} \text{ since } P\alpha VI \tag{7.15}$$

where I_1 in this example is the current at 15 kV, A_1 the conductor cross-section and I_2 is the current at 220 kV and A_2 the corresponding size of the conductor. ℓ designates a certain length and ρ the resistivity of the conductor.

We can see from Eq. (7.15) that for a given percentage of energy loss, the cross-section (and consequently the weight) of the conductors to transmit a given amount of power varies inversely as the square of the

[1] Clearly, the number of poles P needed is $(2\nu/n)$ where n is the rotor speed and ν the frequency of the ac desired. (one electromagnet has two poles, north and south). For example, if the hydrogenerator turbine rotates at 600 rpm or 10 rps (rotations per second) as mentioned above, and a 50 Hz or 50 rps ac supply is needed, the number of poles is $(2 \times 50/10) = 10$.

voltage.[2] The diameter of the copper conductor necessary to transmit 210 MW of power at 220 kV would be about 0.57 cm instead of 8 cm at 15 kV weighing 150 kg/km.

We can conclude from the above discussion that the voltage of generated power should be *increased* (or stepped up) for transmission involving minimum power loss, minimum voltage drop at least cost and again *decreased* (or stepped down) to lower voltages for safe industrial and domestic use. In brief, the voltage needs to be suitably *transformed* and the device which can transform ac voltages and currents is known as *transformer*.

A transformer consists essentially of two sets of coils, insulated from each other, and wound around a common iron core, somewhat as shown in Fig. 7.13. One of the coils is called the *primary* winding and the other is called the *secondary* winding. The number of turns in these windings is different. Let us suppose the primary winding has N_p turns and the secondary winding has N_s turns. If an ac source of emf is connected across the primary winding, we shall show below that the emfs across the primary and the secondary windings which are $\mathcal{E}_p$ and $\mathcal{E}_s$ respectively are related as

$$\frac{\mathcal{E}_s}{\mathcal{E}_p} = \frac{N_s}{N_p} \text{ or } \mathcal{E}_s = \mathcal{E}_p(N_s/N_p). \qquad (7.16)$$

If $N_s > N_p$, that is the secondary has more turns than the primary, $\mathcal{E}_s > \mathcal{E}_p$. The primary voltage is *increased* and the transformer is a *step up* transformer.

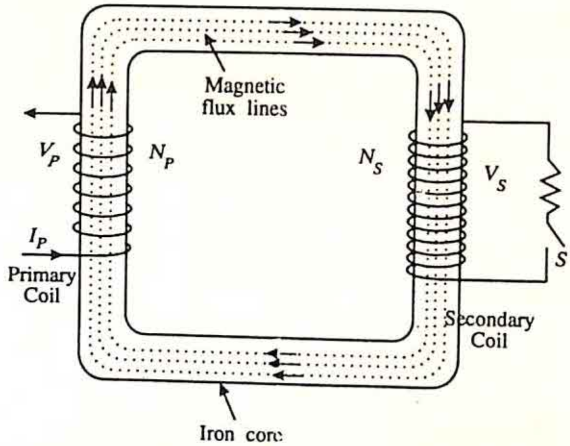


Figure 7.13: Schematic diagram for a transformer, showing a primary coil (N_p turns, carrying an ac current I_p with voltage V_p across the coil) and a secondary coil (N_s turns, voltage V_s, current I_s if connected). Both primary and secondary are wound round a common iron core.

If, on the other hand $N_s < N_p$, $\mathcal{E}_s < \mathcal{E}_p$ and the primary voltage is reduced and the transformer is a *step down* transformer.

We now show that Eq.(7.16) is a direct consequence of Faraday's law of electromagnetic induction. Indeed, we have come across a form of this already in our description of Faraday's experiments, Fig. 7.1. Let an alternating voltage V_p be applied to the primary winding. This causes an alternating current $I_p \cos \omega t$, of angular frequency ω to flow through the primary. Because of this current, a magnetic field is created and a magnetic flux $\phi \cos \omega t$ passes through each turn of the primary coil. Then, from Faraday's law of induction, the emf across the ends of the N_p primary windings is

$$\begin{aligned} \mathcal{E}_p &= -\frac{\mathrm{d}}{\mathrm{d}t}(N_p \phi \cos \omega t) \\ &= \omega N_p \phi \sin \omega t. \end{aligned} \qquad (7.17)$$

If we take an ideal transformer in which all

[2] In actual practice the calculation of conductor size is not as simple as given in Eq. (7.15). There are various other electrical and mechanical considerations that need to be taken into account. The capacity of a conductor to dissipate heat and its tensile strength, are important considerations. Whether conductors run overhead or are insulated as cables and run underground are major factors to be taken into account in designing a conductor. Copper conductors are seldom used now-a-days for overhead transmission of power.

the magnetic flux lines produced by the primary link the secondary winding and the resistances of the windings are negligibly small, the applied voltage V_p is equal and opposite the induced voltage $\mathcal{E}_p$.

Now consider the secondary. Since the secondary is wound such that the *same* flux $\phi \cos \omega t$ passes through each turn of it, the emf across *its* terminals is

$$\begin{aligned}\mathcal{E}_s &= -\frac{\mathrm{d}}{\mathrm{d}t}(N_s \phi \cos \omega t) \\ &= N_s \omega \phi \sin \omega t \qquad (7.18)\end{aligned}$$

Comparing Eqns. (7.18) and (7.17), we see that

$$(\mathcal{E}_s/\mathcal{E}_p) = (N_s/N_p). \qquad (7.19)$$

The ratio of the open circuit emf $\mathcal{E}_s$ in the secondary to the emf $\mathcal{E}_p$ in the primary is equal to the ratio (N_s/N_p) of the number of turns.

Two important points need to be noted:

(i) A transformer operates only for time varying ac voltages and would not operate for dc voltages

(ii) It is the different number of turns in the primary and secondary winding which results in the transformation.

Let us now examine what happens when the switch S in Fig. 7.13 is closed. A current I_s flows through the resistor (which may represent a set of bulbs) and the secondary delivers power $P = I_s^2 R. = \mathcal{E}_s I_s$. Where does this power come from? The power and the energy obviously come from the source (falling water or burning coal), through the turbine, the generator and the primary winding. As soon as a current I_s flows in the secondary winding a corresponding current I_p flows into the primary winding so that the magnitude of power $P = \mathcal{E}_s I_s = \mathcal{E}_p I_p$, if we assume that no losses take place in the transformer. As soon as

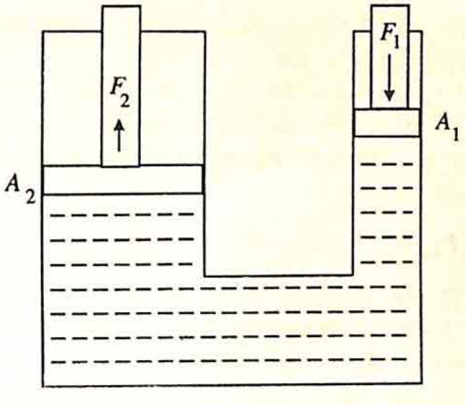


Figure 7.14: Hydraulic analogy for the transformer.

the primary winding draws the power, the generator tends to slow down under the load and a device called a *governor* immediately comes into action letting in more steam (or water) into the turbine so that the generator speed remains unaltered. A very interesting equation which follows from the relation $P = \mathcal{E}_s I_s = \mathcal{E}_p I_p$ is that

$$\frac{\mathcal{E}_s}{\mathcal{E}_p} = \frac{I_p}{I_s} = \frac{N_s}{N_p} \qquad (7.20)$$

This relates currents in the primary and secondary of the transformer.[3] We also note that the transformer windings should have a very low electrical resistance so that there is no power loss when the voltage is transformed. This is always done, but some joule heating and loss are inevitable, so that all the power supplied by the primary is not converted to useable electric power in the secondary. To carry away the heat produced, the iron core with the primary and the secondary windings are immersed in oil which circulates through tubes or fins to dissipate the heat.

We conclude by giving a simple analogy

[3]It is only the magnitudes of the primary and the secondary current that are related as in Eq. (7.20). Their directions are opposite to each other.

which will help us to understand the functioning of a transformer (Fig. 7.14). You have read about the hydraulic press in which a small force F_1 applied at the piston of area A_1 is amplified to F_2 at the piston of area A_2 where

$$(F_2/F_1) = (A_2/A_1) \tag{7.21}$$

Since the pressure p is the same throughout a fluid (neglecting gravity), and force = (pressure × area), we have

$$p = (F_1/A_1) = (F_2/A_2) \tag{7.22}$$

which proves the relation, Eq. (7.21). In an analogous way, the N_p turns of the primary and N_s turns of the secondary winding link the same magnetic flux ϕ and

$$\frac{\mathcal{E}_s}{N_s} = \frac{\mathcal{E}_p}{N_p} = -\frac{\mathrm{d}\phi}{\mathrm{d}t}$$

according to Faraday-Lenz's law.

A hydraulic press is used to lift very heavy objects. You may have seen motor cars being lifted up in a service station. This is done with a hydraulic press. The fact that a small force is amplified many times does not in any way violate the principle of energy conservation. The work done in pushing the piston A_1 through x_1 is F_1x_1 and this must be the same as F_2x_2, the work done in lifting the piston A_2. Since $F_2 >> F_1, x_2 << x_1$. In fact

$$\frac{F_2}{F_1} = \frac{A_2}{A_1} = \frac{x_1}{x_2} \tag{7.23}$$

In a transformer also energy input is equal to the energy output if the losses in the transformer are ignored.

Example 7.5: A 10kVA transformer has 20 turns in the primary and 100 turns in the secondary circuit. An ac voltage $V_p = 600 \sin 314t$ is applied to the primary. Find

(i) the maximum value of the flux

(ii) the maximum value of the secondary voltage.

Answer: We have seen (Eq. (7.17)) that

$$V_p = V_m \sin \omega t \quad \text{where} \quad V_m = N_p \omega \phi$$

V_m being the maximum or peak value of the voltage and $\omega = 2\pi\nu$ = angular frequency in rad/s.

If $\omega = 314, \nu = (314/2\pi) = 50$ Hz. V_m in the present problem is 600 V.

(i) The peak value of flux $\phi = 600/(314 \times 20) = 0.0955$ Wb.

(ii) $\mathcal{E}_s = (N_s/N_p) \times 600 = (100/20) \times 600 = 3000$ V.

Three phase ac power at 11 kV is used for distribution of electric power as you will learn in Section 7.10. It is then stepped down further and we receive single phase power at 240 V for domestic use. Let us imagine that we have to supply 200 kW of ac power to a village which is 15 km away from a substation. Let us try to calculate the losses and the voltage drop if we were to supply the power at single phase at: (a) 240V and (b) 11kV and we shall understand why we cannot use low voltage for transporting the power. (We shall assume that the line has only resistance and no capacitance or inductance (Chapter 8) which are very important when we deal with ac voltages and currents. We shall also assume that the 200 kW load is only due to lighting load made of resistors only). The resistance of the line is 0.01 Ω/km.

The current I_1, to transport 200kW at 240V from the substation would be

$$I_1 = \frac{200,000}{240} = 833.3 \text{ A}$$

The resistance of the line is

$$R = 0.01 \times 15 = 0.15\Omega.$$

The voltage drop is

$$RI = 0.15 \times 833.3\,\text{V} = 125\text{V}.$$

This means that the voltage has dropped to $240 - 125 = 115$V by the time the electricity is received at the village. If you have a 100W bulb rated at 240V, it will glow at

$$100 \times \left(\frac{115}{240}\right)^2 = 23\text{W}$$

This is unacceptable (power consumed by an electric bulb or a resistor R is

$$P = V^2/R \text{ or, } P_1/P_2 = (V_1/V_2)^2).$$

How much is the I^2R loss in the line itself? The I^2R loss is

$$\begin{aligned} P_L &= 833.3^2 \times 0.15\text{W} \\ &= 104.16\text{kW}. \end{aligned}$$

More than half of the power is wasted in transporting the power! If we had used *the same line* and transmitted the power at 11 kV, the current would have come down to

$$I_2 = \frac{200,000}{11,000} = 18.18\text{A}.$$

The voltage drop would have been

$$18.18 \times 0.15 = 2.73\ \text{V}$$

which is negligible and the I^2R loss would be

$$18.18^2 \times 0.15 = 49.6\text{W}$$

which is quite low.

It is observed that we have to distribute power to the village at 11 kV and then step it down to 415 V as described in Section 7.10 and it is always done as 3 phase power.

7.10 Generation of three phase power and its use

In large ac generators, the electromagnet is rotated inside three sets of coils and not inside one coil as we have described so far (Fig. 7.15). The coils are so arranged that the voltages induced in all three coils often

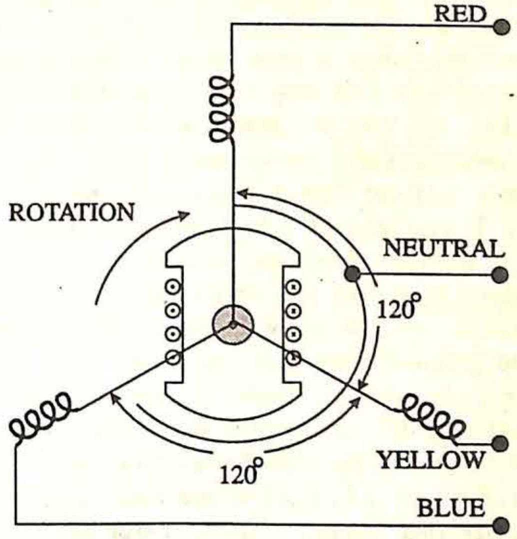


Figure 7.15: Schematic diagram of a 3-phase generator. In a real generator, an electromagnet is rotated, inside sets of coils instead of a coil being rotated in a magnetic field. Three coils (3-phase winding) marked R, Y and B are so placed that their axes are at an angle of 120° with respect to each other. The voltages generated in these coils are also phase displaced by 120° as shown in Fig. 7.16.

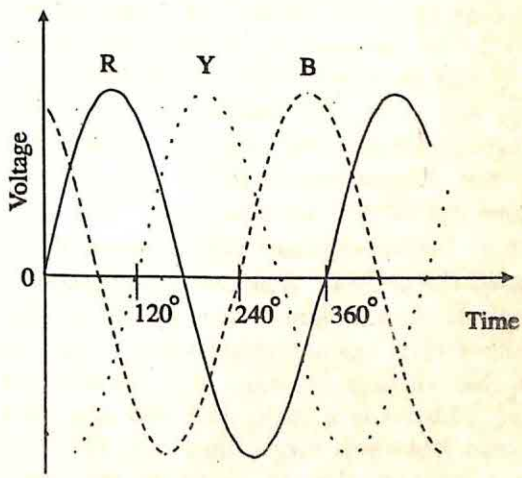


Figure 7.16: Three phase voltages produced in a generator.

described as R (for red), Y (yellow) and B (blue) are the same in magnitude but have their maximum or peak values shifted by angles of 120° with respect to each other (Fig. 7.16). All over the world, electric power is generated and transmitted in three phases since it is convenient and economical to do so. If you observe a high voltage transmission system where huge towers support the conductors, you will notice, in some cases, that there are four conductors. One is called the *ground wire* and the other three are for three phases. Power is generated near coal fields or close to sources of hydroelectric power and transmitted over long distances to cities or industries where large amounts of electrical energy is used. Large amounts of power can only be transmitted over long distances at high voltage as we discussed in the Section 7.9.

In India, the high voltage transmission is at 220 kV and 400 kV. You may have seen big substations where power is received at high voltage and stepped down with the help of *step down transformers* to 110 kV or 66 kV. The voltage is further stepped down to 11 kV for the *distribution of power*. This voltage of 11 kV is still very high for supplying to our homes. If you look around you may see a structure not far from where you live, on which a transformer is mounted. Three phase power at 11 kV, is stepped down by this transformer to 415 volts. Four wires come out of this transformer. Three wires are for the three phases and the fourth wire is called the neutral. When electric connection is given to your house, *one of the phases*, R or Y or B and the neutral are brought in. **If the voltage measured between any two phases is 415 V, the voltage measured between any phase and the neutral wire is $415/\sqrt{3}$ volts = 240 volts.** This voltage is still considered to be high in some countries and for domestic use power is supplied at 110 volts in USA and Canada.

Two pin sockets in your household, where you plug in a radio or a table lamp, are connected to a phase and the neutral. Three pin sockets where you connect a refrigerator or a cooker have three holes. The two lower (and smaller) ones are connected to the phase and the neutral. The top (and the bigger) one is connected to the ground. This is for safety. If the metal body of the refrigerator accidentally gets connected to the phase wire (or the live wire as it is called at times) because of poor connection or due to the insulation wearing out, you would get a nasty electric shock if you touch the refrigerator (the shock will be much worse if you are wet because your body resistance will be low). On the other hand, you may not feel anything if you wear rubber slippers. If the refrigerator is properly grounded electricity will prefer to flow to the earth through the low resistance ground wire rather than through the high resistance of your body. Any electric appliance that has a metal body should be properly grounded.

In India, we have nearly 80,000 MW of power producing capacity out of which about 20,000 MW of power is from hydroelectric sources and about 55,000 MW from coal fired thermal stations. We generate only about 3000 MW from nuclear power stations.

Most of our 5,60,000 villages are now electrified and more than 9 million pumpsets run by electric motors are used for irrigation.

Electric generators, as you must have gathered by now, convert mechanical energy (used for rotating electromagnets) into electrical energy. Electric motors do just the opposite. They convert electrical energy into mechanical energy. Most of the ac motors are induction motors. These motors are usually run from a 3-phase supply. But small motors, like the ones that are used in our

ceiling fans are adapted to run from single phase supply which is the normal domestic power supply comprising one of the 3-phases, and the neutral as we have just mentioned.

The simple experiments carried out by Michael Faraday and Joseph Henry on electromagnetic induction, more than 150 years ago, have transformed this world. What started with an attempt to establish a relationship between magnetism and electricity has resulted in the generation of many trillions of units of electricity all over the world, lighting up this globe where it gets dark with sunset, running industries, running trains. pumping water to irrigate lands, helping to link the world through satellite communication, running computers and doing virtually everything for us with great efficiency. Yet, all that is involved in the generation of electricity is to rotate an electromagnet inside a set of coils as we have seen earlier.

Summary

1. *Faraday's experiments* demonstrated that an emf is induced in a coil if magnetic flux through the coil changes with time. The change in magnetic flux may be brought about, for examples, by suddenly moving a magnet towards (or away) from the coil or during the short time a current is switched on or off in an adjacent coil. The phenomenon is known as electromagnetic induction.

2. An emf is induced between the ends of a wire or between the sides of a closed loop as it moves across a stationary magnetic field. Although the ends of the loop may be joined no current flows as long as the loop is wholly inside the field. In this situation, the two sides of the loop are like two cells opposing the emfs of each other. When the moving loop is partially outside and partially inside the field, a current flows, as when the two terminals are connected by an external conductor. All this can be readily understood in terms of the magnetic Lorentz force on electrons in conductors moving across a magnetic field.

3. *Faraday's law* states that the induced emf $\mathcal{E}$ in a circuit is proportional to the time rate of change of magnetic flux Φ through the circuit or linked by the circuit:

$$\mathcal{E} = -N\frac{\mathrm{d}\Phi}{\mathrm{d}t}$$

The negative sign is indicative of the direction of the induced emf, which is best expressed in terms of *Lenz's law* (or rule): *the direction of the induced current is such that it opposes the flux changes that cause it.* Lenz's rule, in effect, states that the phenomenon of electromagnetic induction is consistent with the principle of energy conservation.

Faraday's law is applicable irrespective of the mechanism that causes the flux change, whether on account of the magnetic field changing across a stationary loop or on account of the loop moving in a steady magnetic field.

4. *An a.c. generator* basically consists of a loop of wire rotating in the field of a permanent magnet. If it rotates with angular speed ω in a field $\mathbf{B}$, the flux Φ through it varies as

$$\Phi = N\ A\ B \cos\omega t$$

where N is the number of turns in the coil having the area A. (At $t = 0$, the loop is normal to the field). The induced emf $\mathcal{E}$ is given by

$$\mathcal{E} = NAB\omega \sin\omega t$$

and the ac current I through a load of resistance R is

$$I = \frac{NAB\omega}{R}\ \sin\omega t$$

In reality an electromagnet is rotated inside a set of coils. The power dissipated in the load is supplied by the agent rotating the electromagnet. In a hydroelectric power plant, the kinetic energy of falling water from a height is converted into the rotational energy of the turbines connected to the loops of wires in the generator. In a thermal power station, the turbines are rotated by steam produced by boiling water using coal or oil or nuclear matter.

5. *A transformer* consists of two coils (Primary & Secondary) wound around a closed magnetic core. If an alternating current flows through the primary coil, an induced emf is set up across the terminals of the secondary coil. If the field lines are confined to the core, the fluxes linked by the secondary and primary coils are in proportion to the number of turns in them. The emf across the secondary coil ($\mathcal{E}_s$) and the emf across the primary coil ($\mathcal{E}_p$) are then related by

$$\frac{\mathcal{E}_s}{\mathcal{E}_p} = \frac{N_s}{N_p}$$

By adjusting the 'turns ratio' N_s/N_p, the voltage across the secondary coil can be 'stepped up' or 'stepped down'.

Transformers are used in transmission of electric power. Power is transmitted (from the power station to the sub-stations near the users) at very high voltages to reduce cost and reduce losses. At a sub-station, a step-down transformer converts the high voltage to a lower value. Power is then distributed through feeders and stepped down again to a lower value suitable for users. All over the world, electric power is generated and transmitted in three phases since it is convenient and economical to do so.

Exercises

7.1 Describe with the help of simple experimental set-ups the phenomenon of electromagnetic induction.

7.2 State Faraday's law of electromagnetic induction. What is the significance of the negative sign appearing in the law?

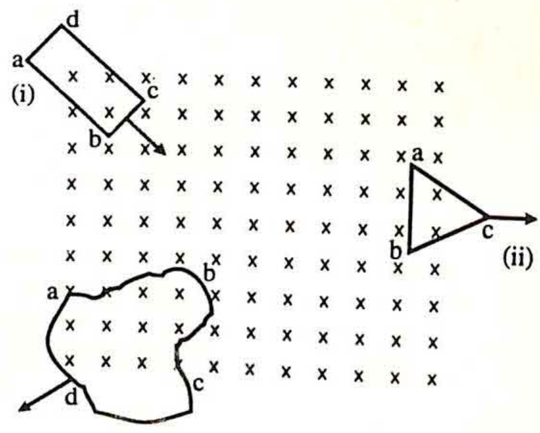
7.3 A small piece of metal wire is dragged across the gap between the pole pieces of a magnet in 0.5s. The magnetic flux between the pole pieces is known to be 8 x 10^{-4} Wb. Estimate the emf induced in the wire.

7.4 A long solenoid with 15 turns per cm has a small loop of area 2.0 cm^2 placed inside normal to the axis of the solenoid. If the current carried by the solenoid changes steadily from 2A to 4A in 0.1 s; what is the induced voltage in the loop while the current is changing?

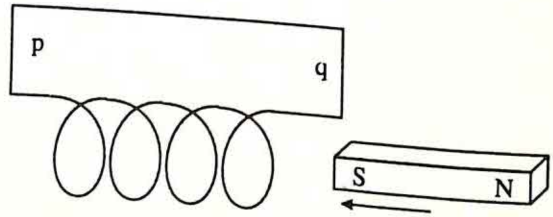
7.5 A 1m long conducting rod rotates with an angular frequency of 400 s^{-1} about an axis normal to the rod passing through its one end. The other end of the rod is in contact with a circular metallic ring. A constant magnetic field of 0.5T parallel to the axis exists everywhere. Calculate the emf developed between the centre and the ring.

7.6 State Lenz's rule for determining the sense of induced current. Explain how the rule is consistent with the principle of conservation of energy.

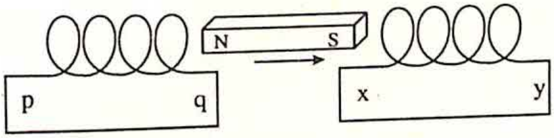
7.7 Figures below show planar loops of different shapes moving out of or into a region of magnetic field which is directed normal to the plane of the loops away from the reader. Determine the direction of induced current in each loop using Lenz's law.



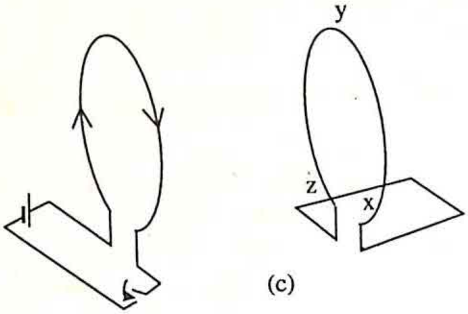
7.8 Predict the direction of induced current in the situations described by the following figures: (a) to (f).



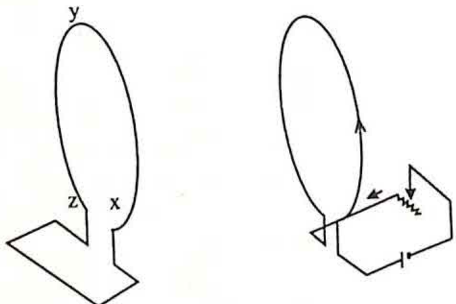
(a)



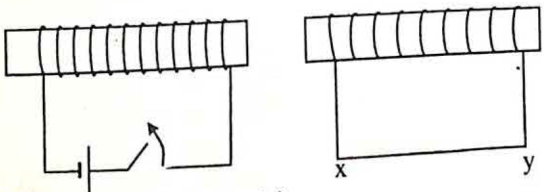
(b)



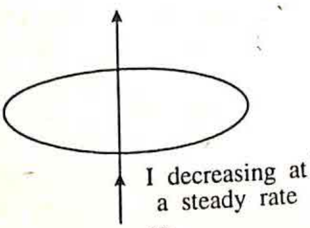
(c)



(d)



(e)



(f)

7.9 Explain with the help of a schematic diagram the principle of an ac generator.

7.10 A circular coil of area 300 cm^2 and 25 turns rotates about its vertical diameter with an angular speed of 40 s^{-1} in a uniform horizontal magnetic field of magnitude 0.05T. Obtain the maximum voltage induced in the coil.

7.11 A circular coil of radius 8.0 cm and 20 turns rotates about its vertical diameter with an angular speed of 50 s^{-1} in a uniform horizontal magnetic field of magnitude 3.0 x 10^{-2} T. Obtain the maximum and average emf induced in the coil. If the coil forms a closed loop of resistance 10Ω, how much power is dissipated as heat? What is the source of this power?

7.12 A wheel with 10 metallic spokes each 0.50 m long is rotated with a speed of 120 rev/min in a plane normal to the earth's magnetic field at the place. If the magnitude of the field is 0.40 G, what is the induced emf between the axle and the rim of the wheel?

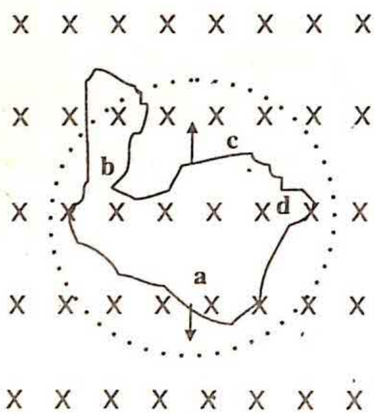
7.13 Explain the use of transformer for long distance transmission of electric power from a power station to actual users.

7.14 A transformer has 300 primary turns and 2400 secondary turns. If the primary supply voltage is 230V, what is the secondary voltage?

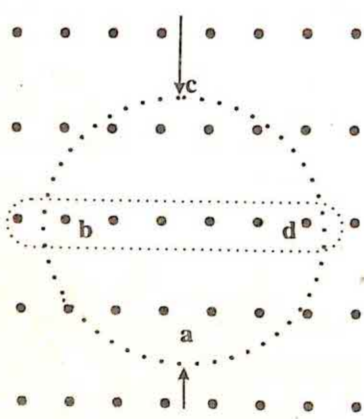
7.15 A transformer has 200 primary turns and 150 secondary turns. If the operating voltage for the load connected to the secondary is measured to be 300 V, what is the voltage supplied to the primary?

Additional Exercises

7.16 Use Lenz's law to determine the direction of induced current in the situations described by the figures below:



(a)



(b)

(a) a wire of irregular shape turning into a circular shape;

(b) a circular loop being deformed into a narrow straight wire.

The cross (×) indicates magnetic field into the paper, and the dot (•) indicates magnetic field out of paper.

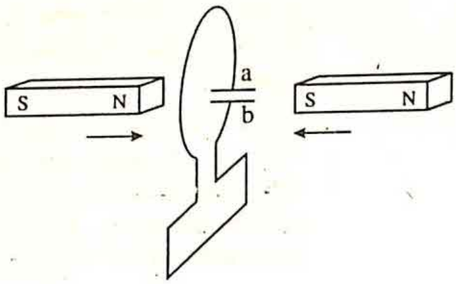
7.17 Answer the following questions:

(a) A conducting loop is held stationary normal to the field between the NS poles of a fixed permanent magnet. By choosing a magnet sufficiently strong, can we hope to generate current in the loop?

(b) A closed conducting loop moves normal to the electric field between the plates of a large capacitor. Is a current induced in the loop when it is (i) wholly inside the capacitor, (ii) partially outside the plates of capacitor? The electric field is normal to the plane of the loop.

(c) A rectangular loop and a circular loop are moving out of a uniform magnetic field region to a field free region with a *constant velocity*. In which loop do you expect the induced emf to be constant *during* the passage out of the field region? The field is normal to the loops.

(d) Predict the polarity of the capacitor in the situation described by the figure below:



7.18 A rectangular loop of sides 8cm and 2cm with a small cut is moving out of a region of uniform magnetic field of magnitude 0.3T directed normal to the loop. What is the voltage developed across the cut if the velocity of the loop is 1cm s^{-1} in a direction normal to the (i) longer side, (ii) shorter side of the loop? For how long does the induced voltage last in each case?

(**Note:** This and some other exercises ignore one important point for simplicity: a magnetic field cannot abruptly change in space from a finite value to zero).

7.19 Suppose the loop in 7.18 is stationary but the current feeding the electromagnet that produces the magnetic field is gradually reduced so that the field decreases from its initial value of 0.3T at the rate of 0.02 Ts^{-1}. If the cut is joined and the loop has a resistance of 1.6 Ω, how much power is dissipated by the loop as heat? What is the source of this power?

7.20 A square loop of side 12cm with its sides parallel to x and y axes is moved with a velocity of 8 cm s^{-1} in the positive x direction in an environment containing a magnetic field in the positive z-direction. The field is neither uniform in space not constant in time. It has a gradient of 10^{-3} Tcm^{-1} along the negative x direction (i.e. it increases by 10^{-3}T per cm as one moves in the -ve x direction), and it is decreasing in time at the rate of 10^{-3}T s^{-1}. Determine the direction and magnitude of the induced current in the loop if its resistance is 4.5 mΩ.

7.21 It is desired to measure the magnitude of field between the poles of a powerful loudspeaker magnet. A small flat search coil of area 2.0 cm^2 with 25 closely wound turns is positioned normal to the field direction, and then quickly snatched out of the field region. (equivalently, one can give it a quick 90^o turn to bring its plane parallel to the field direction). The total charge flown in the coil (measured by a ballistic galvanometer connected to the coil) is 7.5 mC. The resistance of the coil and the galvanometer 0.50 Ω. Estimate the field strength of the magnet.

7.22 Figure shows a metal rod PQ resting on the rails AB and positioned between the poles of a permanent magnet. The rails, the rod and the magnetic field are in three mutually perpendicular directions. A galvanometer G connects the rails through a switch K. Length of the rod = 15cm, B = 0.50 T, resistance of the closed loop containing the rod = 9.0 mΩ.

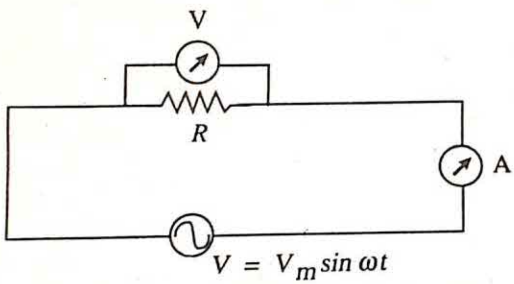


Figure 8.1: An alternating voltage $V(t) = V_m \sin \omega t$ applied to a resistor

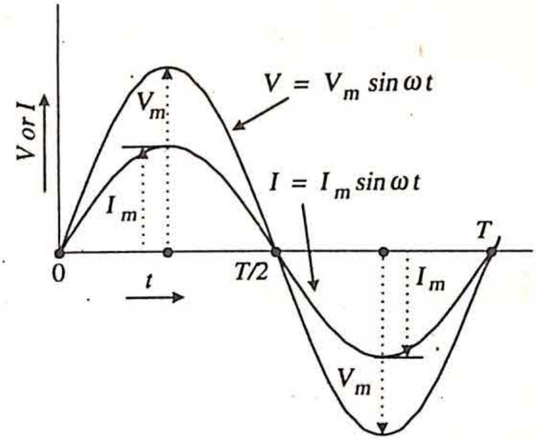


Figure 8.2: The voltage and current in a resistor depend on time in exactly the same way. For an ac supply $V(t) = V_m \sin \omega t$ shown, the electric current is $I(t) = (V_m/R) \sin \omega t$. For example, the zeroes and peaks of the voltage and of the current occur at the same instants of time, as shown above by the dots.

are related to each other in different kinds of circuits. The basic components of such circuits are three, namely resistor (R), capacitor (symbol C) and inductor (symbol L).

In Section 8.2, we consider a resistor across which a voltage varying with time is applied, and introduce the ideas of maximum as well as root mean square (rms) values of voltage and current. We then describe, in Section 8.3, some new effects that arise when a time varying voltage is applied to a capacitor (C) and to an inductor (L). The inductor, a new circuit element, is discussed in Section 8.4. The next few sections (8.5 to 8.8) analyze quantitatively the behaviour of circuits with inductor, capacitor and resistor in different combinations. Among other things, it is found that an inductor and a capacitor (in series or in parallel) have a natural frequency for electrical oscillations. Such resonant circuits (discussed in Section 8.9) are at the heart, for example, of transmitters and receivers of radio waves, which are discussed (along with other electromagnetic waves) in the next chapter (Chapter 9).

8.2 AC voltage applied to a resistor

When a steady current I passes through a resistor R, the voltage drop V across the resistor is

$$V = RI \qquad (8.1)$$

This is Ohm's law, discussed first in Section 3.3. It is found experimentally that the current through a resistor *depends on time*, exactly as the voltage across it, namely

$$V(t) = RI(t) \qquad (8.2)$$

where $I(t)$ is a time varying current. An example is a voltage which alternates sinusoidally with time, i.e. an ac voltage with angular frequency ω, so that

$$V(t) = V_m \sin \omega t \qquad (8.3)$$

In this case (Fig. 8.1), the Ohm's law, Eq. (8.2) for time varying voltages and currents implies that

$$I(t) = (V(t)/R) = (V_m/R) \sin \omega t. \qquad (8.4)$$

In reality, the resistance R does depend somewhat on the angular frequency ω of the applied voltage. We shall neglect this dependence. The ac voltage and current for such a resistor are shown in Fig. 8.2.

There is a convenient way of showing (representing) graphically, quantities such as $V(t)$ in Eq. (8.3), which have a magnitude (V_m) and a phase angle $\theta(=\omega t)$. These two quantities can be represented by a vector in a plane. The length of the vector represents the magnitude or maximum value (V_m) and the angle of this vector, for example, with the x axis represents the phase θ (Fig. 8.3).

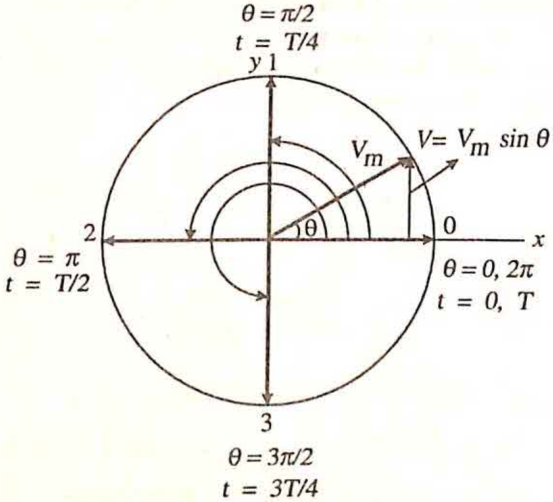


Figure 8.3: The phasor diagram, shown here for a voltage $V(t) = V_m \sin \omega t$. V_m is the length of the vector, shown by a thick line. Its direction or angle θ with respect the x axis represents the phase ωt of the voltage ($\theta = \omega t$). The projection $V_m \sin \theta$ of the vector on the y axis is the actual voltage $V_m \sin \omega t$ at any time. Since the phase angle $\theta = \omega t$, the vector rotates with a uniform angular velocity ω. The figure shows the vector at a general time t, and at four specific times, $t = 0, t = (T/4), t = (T/2)$, and $t = (3T/4)$.

This is known as the *phasor* or *Argand diagram*. Since in the case above, $\theta = \omega t$, the phase angle increases at a uniform rate ω, i.e. the vector rotates counterclockwise with a constant angular velocity ω. The actual voltage for a given phase angle θ is the projection of the vector on the y-axis, as can be seen from Eq. (8.3). As we know from the Chapter on simple harmonic motion (Class XI Physics textbook, Chapter 12, Section 12.9) the projection of a uniformly rotating vector on a diameter executes simple harmonic motion; this is described by Eq. (8.3). We show in Fig. 8.3, the voltage at different times using the phasor diagram.

(0) At $t = 0, \omega t = \theta = 0$ and the phase angle is zero. The vector points along the x direction. Its projection on the y axis vanishes.

(1) At $t = T/4$ where T is the time period of oscillation (related to angular velocity ω by $T = (2\pi/\omega)$, and to frequency ν by $T = (1/\nu)), \theta = \omega T/4 = \omega \times (2\pi/\omega) \times 1/4 = \pi/2$. The position of the vector is along the positive y axis, and the projection on y axis is V_m.

(2) At $t = (T/2), \theta = \pi$ and the vector points along the negative x axis, with zero projection on y-axis.

(3) At $t = (3T/4)$ or $\theta = (3\pi/2)$, the voltage is seen to attain its largest negative value $-V_m$.

(4) Finally, at $t = T$, $\theta = 2\pi$, and the voltage is the same as at $T = 0$.

Thus Fig. 8.3 with uniformly rotating vector of length V_m, describes the sinusoidally varying voltage shown in Fig. 8.2. The current I changes in phase with the voltage. This is true for a resistor.

In some cases, the current and voltage may alternate sinusoidally with the same frequency, but the current may lag behind or lead the voltage, depending on the circuit through which the current flows. This is analogous to two cars running at the same speed, with one following the other at a distance. Or even more precisely, it is like two

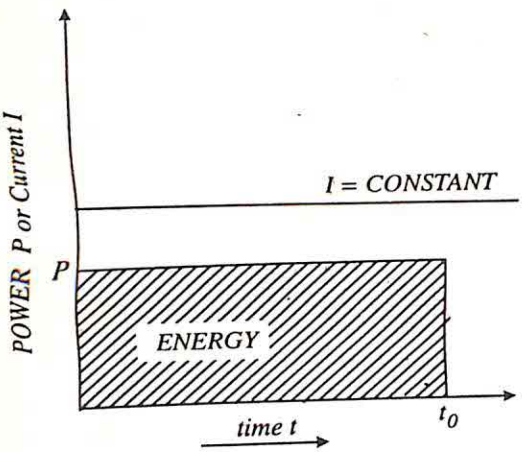


Figure 8.4: A steady or constant current I passing through a resistor R, shown as a function of time t. The power P dissipated ($P = I^2R$) is also steady and is shown. The energy dissipated as heat over a time t_0 is $P\,t_0$, i.e. the area of the shaded rectangle.

pendulums of the same frequency that are started at different times.

For example, if the current I *lags* behind the voltage (which varies as $V_m \sin \omega t$), it can be written as

$$I = I_m \sin(\omega t - \theta). \tag{8.5a}$$

It is written as

$$I = I_m \sin(\omega t + \theta) \tag{8.5b}$$

when the current *leads* the voltage. (In both Eqs (8.5a) and (8.5b), the phase angle θ is positive).

8.2.1 Power loss and rms values

When an electric current I flows through a resistance R, there is power loss of I^2R in the resistor causing heat to be produced. This has been illustrated in Chapter 7, Section 7.9.

In a circuit carrying ac, although the current keeps changing its direction, power loss takes place irrespective of whether the current is positive or negative, since $(+I)^2R$ and $(-I)^2R$ are both positive quantities. The heat energy produced is equal to I^2Rt_0 when a steady current I flows through R for time t_0. This is represented in Fig. 8.4 where the shaded area represents the heat energy produced when a steady current (i.e. $I^2R = P$ = const.) flows for time t_0. When an alternating current I flows through the resistance R, the current does not remain constant with time and therefore the power loss changes from instant to instant.

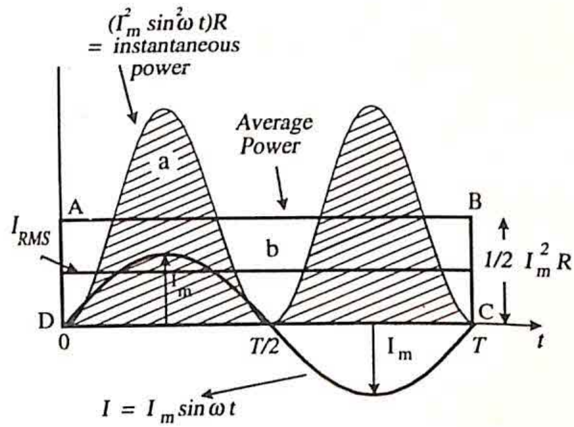


Figure 8.5: An alternating current shown as a function of time t ($I(t) = I_m \sin \omega t$). The power dissipated, namely $P = V(t)I(t) = I_m^2 R \sin^2 \omega t$, is also shown as a function of time, as the shaded area. This area ($I_m^2RT/2$) is equal to that of the rectangle ABCD which is the energy dissipated in a period T when a constant current ($I_m/\sqrt{2}$) flows through the resistor R.

$$\begin{aligned} P &= V(t)I(t) \\ &= (I_m \sin \omega t)^2 R \\ &= I_m^2 R \sin^2 \omega t \\ &= \frac{I_m^2 R}{2}(1 - \cos 2\omega t). \end{aligned} \tag{8.6}$$

The instantaneous power can therefore be

represented by a constant power $I_m^2 R/2$ (the line AB in Fig. 8.5) and an alternating component $(I_m^2 R/2)\cos 2\omega t$ subtracted from it. Over a complete cycle $(I_m^2 R/2)\cos 2\omega t$ cancels out (for example the area a cancels area b). Over a cycle of period the total energy loss is $[(I_m^2 R)/2].\ T = \left(I_m/\sqrt{2}\right)^2 R\, T$ as if a direct current $\left(I_m/\sqrt{2}\right)$ has been flowing in the circuit producing $\left(I_m/\sqrt{2}\right)^2 R\, T$ energy loss. The average power loss is

$$P_{av} = \frac{1}{T}\left[\left(\frac{I_m}{\sqrt{2}}\right)^2 RT\right] = \left(\frac{I_m}{\sqrt{2}}\right)^2 R$$

$$= I_{\text{rms}}^2 R \tag{8.7}$$

$(I_m/\sqrt{2})$ is called the Root Mean Square value of the current (so called because we take the Square of the current, then the Mean and then the Root of the result). For a sinusoidal voltage $V_m \sin\omega t$, the rms value is $V_m/\sqrt{2}$.

The rms current and voltage are given symbols I_{rms} and V_{rms}, respectively.

Example 8.1: A sinusoidal voltage $V = 200 \sin 314t$ is applied to a resistor of 10Ω resistance. Calculate

a) the frequency of the supply;

b) the rms value of the voltage;

c) the rms value of the current; and

d) the power dissipated as heat in watts.

Answer:

a) Since $\omega = 314$, the frequency $\nu = (\omega/2\pi) = 50\text{Hz}$.

b) $V = V_m \sin\omega t = 200 \sin 314t$

$$V_{\text{rms}} = \frac{V_m}{\sqrt{2}} = \frac{200}{\sqrt{2}} = 141.4 \text{ V}$$

c) $$I_{\text{rms}} = \frac{V_{\text{rms}}}{R} = \frac{141.4}{10} = 14.14 \text{ A}.$$

d) The power dissipated is P, where

$$P = \frac{I_m^2}{2} R = I_{\text{rms}}^2 R$$
$$= (14.14)^2 \times 10 = 2000 \text{ W}$$
$$P = \frac{V_m^2}{2R} = \frac{V_{\text{rms}}^2}{R} = \frac{(141.4)^2}{10}$$
$$= 2000 \text{ W}.$$

8.3 Experimental observations of time varying voltages

We now describe what one would observe with a time varying voltage source applied not just to a resistor, but to a resistor R in series with a capacitor C or an inductor L. The inductor is a new circuit element we shall discuss later in Section 8.4. The time varying source could be a voltage that is turned on or off, it could also be an ac voltage. These observations lead to the conclusion that the capacitor and the inductor also resist or impede the flow of time varying current, just like a resistor impedes the flow of steady current. After a discussion of the inductor in the next section (Section 8.4), we then analyze quantitatively how these circuit elements function (Section 8.5 onwards).

8.3.1 RC series circuit

Consider the simple circuit of Fig. 8.6 consisting of just a resistor R and a capacitor C in series. There is a dc voltage source, and a switch S that makes or breaks the circuit. An ammeter A measures the current through the circuit, and the voltmeters V_R, V_C and V_S measure voltages across the resistor, the capacitor and the source or battery. All these are dc measuring instruments. As soon as you close the switch, the needle of the ammeter A will flick for a moment, and come back to its original position, indicating

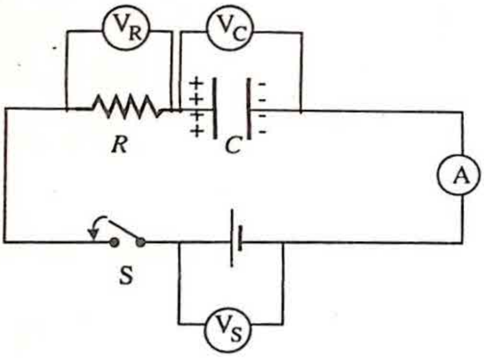


Figure 8.6: A dc voltage is switched on in an RC circuit, i.e. a dc voltage is applied to an RC series circuit by closing the switch S. The needle of the ammeter A initially shows a large deflection which slowly comes back to zero. The voltmeter V_R behaves exactly the same way. The voltmeter V_C initially gives zero reading but ultimately reads the same as V_S.

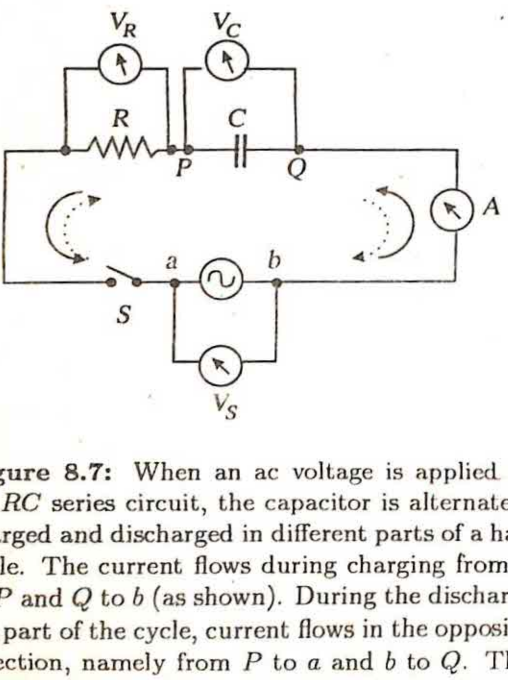


Figure 8.7: When an ac voltage is applied to an RC series circuit, the capacitor is alternately charged and discharged in different parts of a half cycle. The current flows during charging from a to P and Q to b (as shown). During the discharging part of the cycle, current flows in the opposite direction, namely from P to a and b to Q. The ac ammeter and all the ac voltmeters give steady readings. The voltages are related as described in the text.

that a current flows momentarily, and stops flowing thereafter. Although the switch S is closed, no current flows, as if the circuit is open. A capacitor in series results in an open circuit for a dc supply.

You will find it very interesting to observe how the voltmeters behave. Before you close the switch, V_S measures the open circuit voltage V (or emf) of the battery; V_C measures zero voltage. As soon as you close the switch,

(i) V_S moves a little, for just as long as the current flows,

(ii) the voltage V_C across the capacitor, which showed zero, builds up until it shows the same reading as V_S. The capacitor has charged up, with the polarities as shown in Fig. 8.6 to the voltage of the source.

(iii) The voltmeter V_R behaves exactly like the ammeter A, and finally reads zero. No changes take place thereafter.

Now what does this all mean? Clearly, during the time that the voltage V is switched on from zero to V, a current *does* flow in the circuit, though it contains a capacitor. So whereas no *steady* current can flow through a capacitor, it is as if a time dependent current can! We know that the charge Q on a capacitor C is related to the voltage V across it by the equation $Q = CV$. So if the voltage changes as a function of time, the charge also will; this is the current observed during switching. We discuss this in more detail in Section 8.5.

Suppose we carry out the same experiment with an ac voltage source replacing the battery. The dc voltmeters and ammeters have to be replaced by ac instruments. (Such experiments are best done with an ac supply stepped down from the mains. The voltage of 220 V is much too high and dangerous

for using without your teacher's guidance). As soon as you close the switch, the needles of the meters will flick initially, but the final current will no longer be zero. In fact a steady alternating current flows in the circuit. Does the capacitor allow alternating current to flow while it blocks direct current? Not really. The current in the circuit (Fig. 8.7) flows like a see-saw, once from a to P, and then from Q to b, as the capacitor charges and discharges alternately at the same frequency as the source voltage.

Note that the voltage V_R and the voltage V_C do not add up to the supply voltage V. You will see for reasons explained later, that

$$V^2 = V_R^2 + V_C^2 \tag{8.8}$$

just like in the theorem of Pythagoras!. (This can easily be checked by changing R or C or both, while keeping V fixed). Also, you will see that as the voltage V increases, the current I increases, proportionately, for any given R and C. So, one has a new kind of Ohm's law

$$V \propto I \qquad \text{or } V = ZI \tag{8.9}$$

where the constant of proportionality Z is called *impedance*, since it describes how a resistor and a capacitor combine to impede the flow of alternating current. It is not equal to either R or C. We shall relate the impedance to R, C and the ac frequency ω later in Section 8.5.3.

8.3.2 LR series circuit

Let us now consider another type of circuit element, called an inductor. For example, a coil of very high conductivity wire, wound around a core (usually iron) constitutes an inductor. An ideal inductor, which has *no* dc resistance, is connected in series with a resistor, a dc source and a switch as in Fig. 8.8, with a dc ammeter and voltmeters connected as shown. When we close the switch

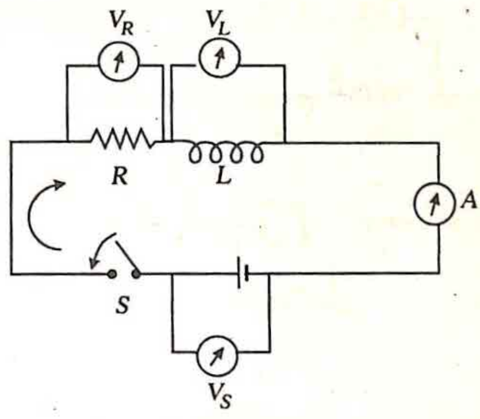


Figure 8.8: When a dc voltage is applied to an LR series circuit, the deflection of the needle of the ammeter A and of the voltmeter V_R grows slowly to its final steady value. Initially V_L shows a large deflection. This becomes zero in the steady state when $V_R \rightarrow V_S$.

the voltmeter V_L shows a voltage instantaneously, which then becomes zero. The needle of the ammeter starts from the zero position and finally reaches a steady value. The voltmeter V_R behaves exactly as the current, growing from zero to a final steady value. If the inductor has no resistance, V_R would finally equal V_S, the supply voltage and the current would, according to Ohm's law, become

$$I = (V_R/R) = (V_S/R).$$

If V_L is zero, it means that the inductor behaves like a short circuit, a commonly used term meaning that the steady current flowing through it experiences no resistance or impedance and that the voltage drop across it is zero.

The question here is, why does V_L assume a nonzero value instantaneouly, and then become zero? The answer lies in Faraday's law. As the current is switched on and the current through the inductor changes with time, so

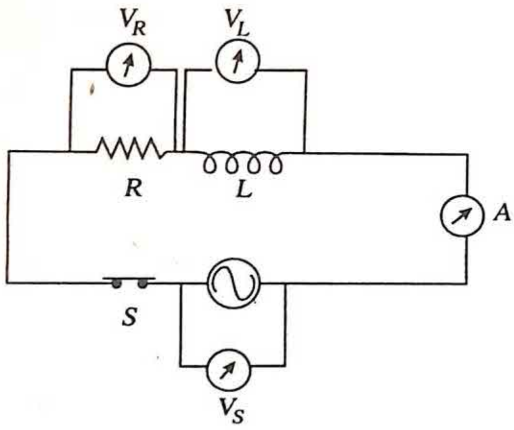


Figure 8.9: An LR circuit to which an ac voltage is applied. It is found that the ac ammeter, and the voltmeters give steady readings, and that $V^2 = V_L^2 + V_R^2$.

does the magnetic flux through the inductor coils. This induces an emf across the ends of the inductor. Suppose the total magnetic flux linked by the inductor is Φ. This is expected to be proportional to the current I flowing through it since this causes the magnetic flux; we have

$$\Phi \alpha\ I \tag{8.10a}$$

or

$$\Phi = L\ I. \tag{8.10b}$$

The constant of proportionality L is called the *self inductance* of the inductor (see detailed discussion in the next section). Clearly, the induced emf $-(d\Phi)/dt)$ is nonzero so long as the current I changes with time, since

$$V_L \simeq \mathcal{E}_L = -(d\Phi/dt) = -L(dI/dt) \tag{8.11}$$

($V_L = \mathcal{E}_L$ for an ideal inductor with no dc resistance).

So because of the self inductance, an opposing emf is set up; it opposes the change in current. This is somewhat like inertia which opposes any change in the state of motion. When there is a steady current, this emf or voltage across the inductor goes to zero.

Suppose now an alternating voltage source replaces the dc source (and ac measuring instruments are used). This is schematically shown in Fig. 8.9. Now, since the alternating current changes continuously in magnitude and direction, there is always a voltage drop V_L across the inductor. Again we find that V_L and V_R obey the Pythagorean relation

$$V^2 = V_L^2 + V_R^2 \tag{8.12a}$$

and that

$$(V/I) = Z \tag{8.12b}$$

is a constant for a given L, R and ac frequency ω. Thus, while an ideal inductor has *zero* resistance for direct current, it has a nonzero impedance for alternating or time varying current (Section 8.6 has more details).

In the next sections, we study these new properties of a capacitor and an inductor and find that they give rise to many effects and applications which you have actually come across. Before that, we briefly discuss the new circuit element mentioned above, namely the inductor.

8.4 Inductance

8.4.1 Self Inductance

Consider a simple solenoid which has been discussed in Chapter 5 (Section 5.4.2). Fig. 5.16 shows the magnetic field inside a solenoid. We found there that if the solenoid carries a current I, and has n turns per unit length, the magnitude of the magnetic field B is given by

$$B = \mu_0 n I$$

If the solenoid is filled with a medium of permeability μ (Section 6.4), the magnetic field is

$$B = \mu n I \tag{8.13}$$

where we have rewritten n as (N/ℓ). N is the total number of turns and ℓ is the actual length of the solenoid. If the solenoid has a cross-sectional area A, *each* turn encloses a flux

$$\Phi = BA = (\mu NA/\ell)I. \tag{8.14}$$

since the magnetic field is perpendicular to the plane of the coil. The total flux linked by the entire solenoid of N turns is

$$\Phi = N\phi = (\mu N^2 A/\ell)I \tag{8.15}$$

Now suppose the current I changes with time. The result, Eq. (8.15), is true at each instant of time. So the magnetic flux linked by the solenoid will change as a result of the current flowing through it, and an emf will be induced. This is an example of *self-inductance*. We see from Eq. (8.11) that

$$\mathcal{E} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\frac{\mu N^2 A}{\ell}\frac{\mathrm{d}I}{\mathrm{d}t} \tag{8.16a}$$

or

$$\mathcal{E} = -L\left(\frac{\mathrm{d}I}{\mathrm{d}t}\right). \tag{8.16b}$$

In this case, i.e. for a solenoid of N turns, length ℓ, cross-sectional area A, and containing a medium of magnetic permeability μ,

$$L = (\mu N^2 A/\ell). \tag{8.17}$$

This relation indicates how a solenoidal inductance of a given L can be made.

In general, a current I flowing through *any* circuit element (not just a solenoid) will cause a magnetic flux Φ to be linked through it. The ratio of the magnetic flux Φ linked by a circuit element to the current I flowing through it is called its *self inductance* L,

$$\Phi = LI. \tag{8.18}$$

Again in general, for any circuit element with self inductance L, the emf $\mathcal{E}_L$ induced between its ends when a current I flows through it is given by

$$\mathcal{E}_L = -L\left(\frac{\mathrm{d}I}{\mathrm{d}t}\right). \tag{8.19}$$

This clearly follows from the definition of self inductance (Eq. (8.18)) and Faraday's law of electromagnetic induction.

The SI unit of inductance is called henry, in honour of Joseph Henry who discovered the phenomenon of electromagnetic induction independently of Faraday. It is given the symbol H. From Eq. (8.19), we see that an inductor has an inductance of 1 henry or 1 H if an emf of one volt develops across its ends when the electric current through it changes at the rate of one ampere per second.

We now show that the energy E stored in an inductor L through which a current $I(t)$ is passing, is $(1/2)LI^2$. If the voltage drop across the inductor is V_L, the power is $V_L I$. But for an inductor, $V_L = L(dI/dt)$, so that power

$$P = V_L I = LI\left(\frac{\mathrm{d}I}{\mathrm{d}t}\right).$$

But power is the rate of change of energy, so that

$$P = \left(\frac{\mathrm{d}E}{\mathrm{d}t}\right) = LI\left(\frac{\mathrm{d}I}{\mathrm{d}t}\right).$$

From this it follows that

$$E = (1/2)LI^2 \tag{8.20}$$

as may be easily checked by differentiating both sides.

8.4.2 Mutual Inductance

Finally, we note that current flowing in one circuit element can produce a magnetic field which may link with another circuit element nearby. So if the current in the first element changes with time, the magnetic flux linked by the second circuit element will also change with time, and thus an emf will be

induced across its ends. This is appropriately enough called *mutual induction.* Suppose a current I flows through a circuit element called 1 and the magnetic flux linked by circuit element 2 due to this current is Φ_2. Then the ratio (Φ_2/I_1) is called the *mutual inductance* M_{12}, i.e.

$$\Phi_2 = M_{12}I_1 \tag{8.21}$$

Again, the emf induced in circuit element 2 is

$$\mathcal{E}_2 = -\frac{\mathrm{d}\Phi_2}{\mathrm{d}t} = -M_{12}\left(\frac{\mathrm{d}I_1}{\mathrm{d}t}\right). \tag{8.22}$$

Though this may look complicated, we have already discussed an example of this, namely the transformer! There the current passing through the primary causes a magnetic field which is linked to the secondary.

Consider, for example, two solenoids S_1 and S_2 one being inside the other (Fig. 8.10). The outer one (S_2) has N_2 turns and radius r_2, while the inner one (S_1) has N_1 turns and radius r_1. The two solenoids have the same length ℓ. Let us find the magnetic flux threading the outer solenoid when a current I_1 flows through the inner solenoid. The magnetic field inside S_1 is $\mu_0 n_1 I_1$ {where $n_1 = (N_1/\ell)$} and that outside S_1 is zero. Thus the magnetic flux through a single turn of S_2 is $(\mu_0 n_1 I_1 \times \pi r_1^2)$ since the flux I_1 outside S_1 is zero. Since there are N_2 turns in S_2, the total flux linked by S_2 is

$$\begin{aligned}\Phi_2 &= N_2(\mu_0 n_1 I_1 \pi r_1^2)\\ &= \mu_0 n_1 n_2 \ell \pi r_1^2 I_1\end{aligned} \tag{8.23}$$

The emf induced in S_2 if the current I_1 in solenoid S_1 changes is given by

$$\mathcal{E}_2 = -\frac{\mathrm{d}\Phi_2}{\mathrm{d}t} = -\mu_0 n_1 n_2 \ell \pi r_1^2 \frac{\mathrm{d}I_1}{\mathrm{d}t}. \tag{8.24}$$

Comparing with Eqs(8.21) and (8.22), which defines mutual inductance, we find that

$$M_{12} = \mu_0(n_1 n_2 \ell \pi r_1^2) \tag{8.25}$$

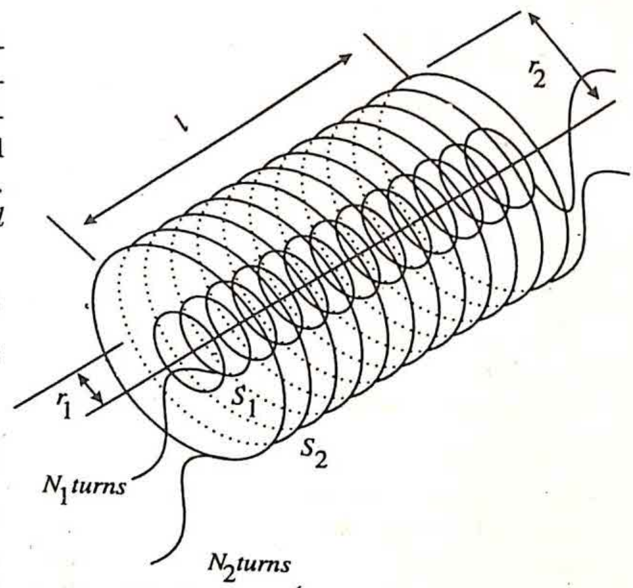


Figure 8.10: Two solenoids, one inside another. The inner solenoid S_1 of radius r_1 and with n_1 turns per unit length is surrounded by the outer solenoid S_2 of radius r_2 with n_2 turns per unit length. Both have the same length ℓ. A time varying current passing through one solenoid induces an emf in the other (mutual inductance).

If the magnetic medium is other than air and has permeability of μ, $M_{12} = \mu(n_1 n_2 \ell \pi r_1^2)$.

It can be explicitly shown, by calculating the flux linked by S_1 due to an electric current in S_2, that

$$M_{21} = M_{12}. \tag{8.26}$$

This is actually a general result.

8.5 Analytical study of the RC circuit

8.5.1 RC series circuit with dc voltage

We now consider in some detail what happens when a dc voltage V is switched on or off in a circuit containing a resistor and a capacitor (Fig 8.6). There are two factors which contribute to the voltage drop V across the circuit. If a current I flows through the resistor R, the voltage drop in

it is IR (reading of the voltmeter V_R) and if there is a charge Q on the capacitor the voltage drop across it is (Q/C) (reading of the voltmeter V_C). So we have

$$RI(t) + \frac{Q(t)}{C} = V(t) \qquad (8.27)$$

where the time dependence of I, Q and V is indicated. Now the rate of change of Q, with time *is* the current, i.e.,

$$\frac{dQ(t)}{dt} = I(t)$$

Substituting this in Eq. (8.27) we have

$$R\frac{dQ}{dt} + \frac{Q}{C} = V$$

$$\text{or} \quad \frac{dQ}{dt} + \frac{Q}{RC} = \frac{V}{R}. \qquad (8.28)$$

We have to solve this equation for Q. We also have to satisfy the condition that initially (at $t = 0$) there is *no* charge on the capacitor, and finally, when V has been switched on, the charge is $Q = CV$.

The correct solution is

$$Q(t) = CV(1 - e^{-t/RC}) \qquad (8.29)$$

You can check this by differentiating and substituting in Eq. (8.28). Also note that this solution satisfies the condition on initial charge and final charge, namely $Q(0) = 0$, and $Q(\infty) = CV$.

The current $I = dQ(t)/dt$ is given by

$$I(t) = (V/R)e^{-t/RC}. \qquad (8.30)$$

The voltage V_C across the capacitor is given by

$$V_C = (Q/C) = V(1 - e^{-t/RC}) \qquad (8.31)$$

The voltage and current are plotted in Fig. 8.11.

The current $I = (V/R)e^{-t/RC}$ jumps from 0 to (V/R) and then decays exponentially to zero as seen in Fig. 8.11a.

The ammeter in the earlier experiment showed a transient current which decayed to

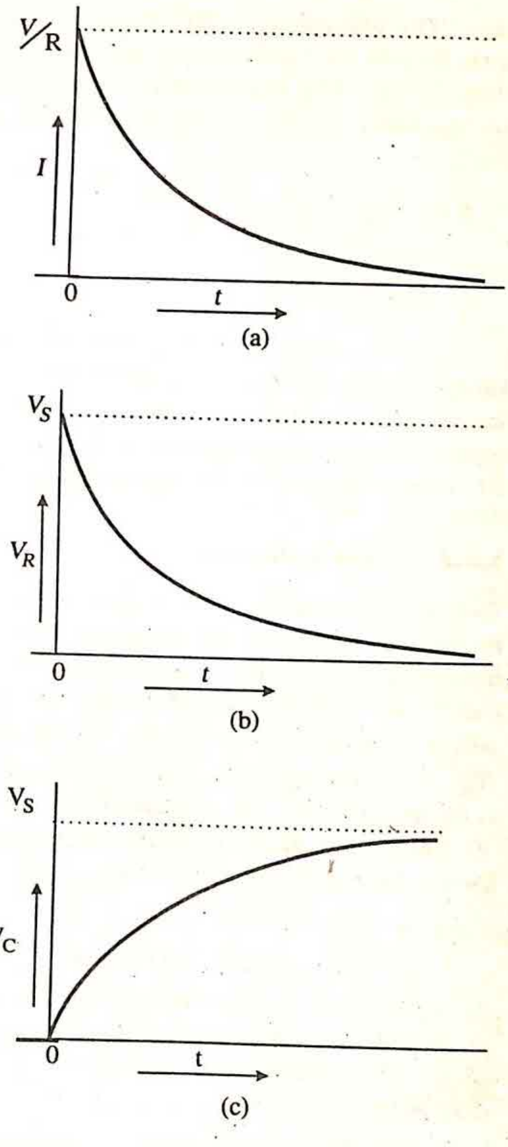


Figure 8.11: The current and voltage in a RC circuit which is being switched on, as in Fig. 8.6. (a) The current I is zero before switching on ($t < 0$), rises immediately to the maximum possible value (V/R), and then decays slowly to zero. V is the source voltage V_S. (b) The reading of the voltmeter across the resistor R. Since $V(t) = RI(t)$, $V(t)$ has exactly the same time dependence as $I(t)$. (c) The voltage V_C across the capacitor saturates to the source voltage V_S.

zero. The voltage $V_R = RI$ (Fig. 8.11b) exactly follows the same pattern as the current (Fig. 8.11a). The charge on the capacitor is zero initially. In Eq. (8.29) if we substitute $t = 0$

$$Q = CV(1 - e^0) = 0$$

and when $t \to \infty$

$$Q = CV(1 - e^{-\infty}) = CV$$

Fig. 8.11c shows how the growth of charge takes place and since $V_C = Q/C$, the reading of the voltmeter V_C (the one across the capacitor) reflects the growth of charge in an RC circuit when a dc voltage is applied.

8.5.2 Time constant

Having said that the current decays quickly or the voltage builds up slowly, it becomes necessary to quantify these terms. We can clearly see that the rate of current decay or voltage rise depends critically on the term RC which is called the *time constant* and is designated as τ. If τ is small the current will decay very quickly and the voltage build up will be rapid. At a time $t = \tau = RC$

$$V_C = V - Ve^{-t/RC} = V(1 - e^{-1})$$
$$= V(1 - 0.3678) = 0.632 \text{ V}$$

So we can define time constant as the time in seconds at which the voltage V_C reaches 63.2% of its final value V (or the current decays to

$$(V/R)e^{-t/RC} = (V/R)e^{-1} = 0.368(V/R)$$

or 36.8% of its initial value). You may verify that RC[1] has the dimension of time.

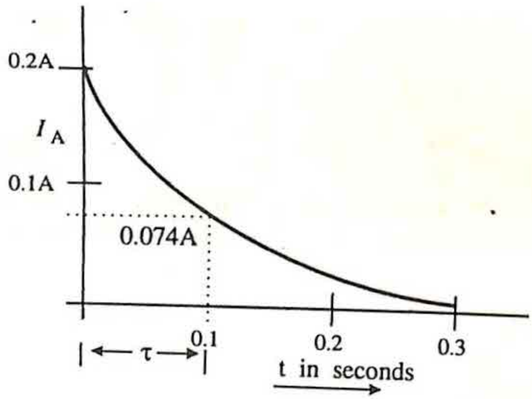
Example 8.2: A dc voltage of 10 volts is applied to an RC series circuit where the resistance is 50 Ω and the capacitance is 2000 μ F. Plot the graphs for the current I and the voltage across the capacitor.

Answer: We proceed in several steps as follows:

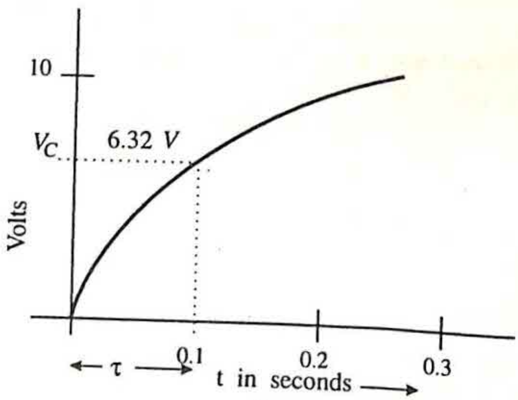
Step 1:

The time constant

$$\tau = RC$$
$$= 50 \times 2000 \times 10^{-6}\text{s}$$
$$= 0.1\text{s}$$



(a)



(b)

Step 2: The current $I(t)$ at any time t is, from Eq. (8.30), given by

$$I = (V/R)e^{-t/(0.1)} = (10/50)e^{-10t}$$

[1] $[R] = [V]/[I]; [C] = [q]/[V]$. $\therefore [R][C] = [q]/[I] = [I][T]/[I] = [T]$

$$= 0.2e^{-10t}\ \text{A}$$

Step 3: The voltage across the capacitor is given by Eq. (8.31), i.e.

$$V_C = V\left(1 - e^{-10t}\right) = 10\left(1 - e^{-10t}\right)\ \text{V}$$

The graphs can be plotted as in Fig. 8.11, and are shown in Fig. (a) and (b).

Note: At $t = 0.1$ s, the time constant of the circuit, V_C has reached 63.2% of its final value and the current has decayed to 36.8% of its initial value, namely, V/R.

8.5.3 RC series circuit with ac voltage

Let us now consider an RC circuit again and analyse the circuit behaviour mathematically when an ac voltage is applied. The voltage equation for the circuit is

$$V_m \sin\omega t = RI + \frac{Q}{C} \tag{8.32a}$$

Q is the charge on the capacitor and the current $I = dQ/dt$.

Therefore,

$$V_m \sin\omega t = R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C}. \tag{8.32b}$$

We are interested to know the current I and charge Q in the circuit as a function of time. A simple way to solve Eq. (8.32) is to *assume* the solution

$$Q = Q_m \sin(\omega t + \phi). \tag{8.33}$$

We are obviously assuming that a sinusoidal voltage will also result in a sinusoidal charge flow having the same frequency. In Eq. (8.33), Q_m stands for the peak value of the charge which we have to find, and the term ϕ stands for the phase relationship between the applied voltage and the charge flowing in the circuit. Using Eq. (8.33) for $Q(t)$ and the corresponding equation

$$\begin{aligned}\frac{\mathrm{d}Q}{\mathrm{d}t} &= \omega Q_m \cos(\omega t + \phi)\\ &= I_m \cos(\omega t + \phi),\end{aligned}$$

Eq. (8.32b) becomes

$$\begin{aligned}V_m \sin\omega t &= \omega I_m[R\cos(\omega t + \phi)\\ &\quad + (1/\omega C)\sin(\omega t + \phi)]\end{aligned} \tag{8.34}$$

Now using standard trigonometry, the quantity in square brackets on the right hand side of Eq. (8.34) can be written as Z cos $(\omega t + \phi - \theta)$ where $Z = \sqrt{R^2 + (1/\omega C)^2}$ and $\tan\theta = (1/\omega CR)$. The steps are as follows:

$$\begin{aligned}&[R\cos(\omega t + \phi) + (1/\omega C)\sin(\omega t + \phi)]\\ &= Z[(R/Z)\cos(\omega t + \phi) +\\ &(X_c/Z)\sin(\omega t + \phi)]\end{aligned} \tag{8.35}$$

where we choose

$$Z = \sqrt{R^2 + X_c^2} \text{ with } X_c = (1/\omega C) \tag{8.36}$$

The quantity $X_c = (1/\omega C)$ is called the *capacitive reactance*. Clearly, if we denote

$$\begin{aligned}&\cos\theta = (R/Z), \sin\theta = (X_c/Z)\\ &\text{or } \tan\theta = (X_c/R),\end{aligned} \tag{8.37}$$

Eq. (8.34) becomes

$$V_m \sin\omega t = \omega Q_m Z\cos(\omega t + \phi - \theta) \tag{8.38}$$

The solution of Eq. (8.38) is

$$V_m = \omega Q_m Z \tag{8.39a}$$

and

$$(\phi - \theta) = -\pi/2 \tag{8.39b}$$

since $\cos(\omega t - \pi/2) = \sin\omega t$. Eq. (8.39) determines Q_m and ϕ in terms of the circuit parameters and hence solves the problem of the behaviour of an RC series circuit when an ac voltage is applied. We now examine this solution.

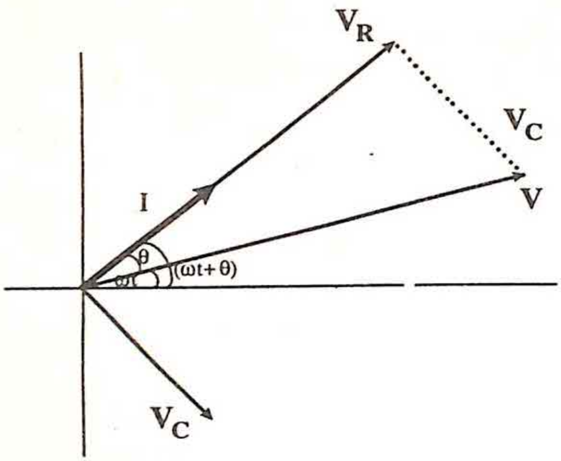


Figure 8.12: Phasor diagram for the RC circuit of Fig. 8.7. The voltage $\mathbf{V}_R$ across the resistor and the current $\mathbf{I}$ through it, the voltage $\mathbf{V}_C$ across the capacitor, and the resultant total voltage $\mathbf{V}$ (supplied by the ac source) are shown as vectors.

The charge Q as a function of the applied voltage is

$$\begin{aligned} Q &= (V_m/\omega Z)\sin(\omega t + \theta - \pi/2) \\ &= -(V_m/\omega Z)\cos(\omega t + \theta) \qquad (8.40) \end{aligned}$$

where Z and θ are given by Eqs (8.36) and (8.37). Now the current is

$$\frac{dQ}{dt} = I = (V_m/Z)\sin(\omega t + \theta) \qquad (8.41)$$

It is clear that the maximum current is (V_m/Z) so that $Z = \sqrt{R^2 + X_c^2}$ is the equivalent resistance or *impedance* of the circuit. Further, the current is *ahead* of the voltage by the angle $\theta = \tan^{-1}(X_c/R)$. *The current is said to lead the voltage.*

8.5.4 Phasor representation of voltages and currents

The voltages and currents in an ac circuit with a resistor and a capacitor in series can be shown graphically using the phasor or vector diagram (Fig. 8.12). We will see that this way one can understand very clearly the meaning of the equations we have written above. We can also find the results for impedance Z, phase angle θ etc, purely geometrically. More complicated ac circuits can also be visualized and analyzed graphically.

You will recall how in Fig. 8.3 the rotating vector V_m generates a sine wave $V = V_m \sin \omega t$. Let us have another sinusoid $I = I_m \sin(\omega t + \phi)$, having the same angular frequency ω but *leading* V by an angle θ. This can then be represented by a vector of magnitude I_m, being ahead of V_m by an angle θ in the counter-clockwise direction (Fig. 8.12). Since V_m and I_m vectors both rotate *with the same angular frequency*, we may 'freeze' them and represent them by stationary vectors with an angle θ separating them. (Since both V and I are sinusoids, their RMS values are $V_m/\sqrt{2}$ and $I_m/\sqrt{2}$ respectively and are often used in place of V_m and I_m which are the maximum values of the voltage and current respectively).

Let us see how the phasor diagram (or the vector diagram) can represent the voltage and currents in the RC circuits that we have just now discussed. Suppose the voltage applied to the circuit be represented by $\mathbf{V}$ in Fig. 8.12. It actually rotates with an angular velocity ω and ωt represents its position at an instant of time t with respect to the x-axis. Since the current *leads* $\mathbf{V}$ by an angle θ, its position is given by $\omega t + \theta$. As has been stated earlier we shall ignore that the vectors actually rotate and consider them as stationary.

The voltage drop across the resistor is $\mathbf{V}_R = R\mathbf{I}$ which is *in phase* with $\mathbf{I}$ and drawn as shown. The voltage across the capactior C is $\mathbf{V}_C$ which *lags* the vector I by an angle of $\pi/2$ and is drawn as shown in Fig. 8.12. (You can see by comparing the expressions for Q in Eq. (8.40) and for I in Eq. (8.41) that Q lags I by an angle of $\pi/2$). The applied voltage V is vector sum of $\mathbf{V}_R$

and $\dot{\mathbf{V}}_C$ and

$$|\mathbf{V}|^2 = |\mathbf{V}_R|^2 + |\mathbf{V}_C|^2$$

as has been stated in Eq. (8.8). We can calculate the phase difference between the applied voltage $\mathbf{V}$ and the current $\mathbf{I}$ as below:

$$\begin{aligned}\tan\theta &= \frac{|\mathbf{V}_C|}{|\mathbf{V}_R|} = \frac{X_C \times I_m}{R \times I_m} \\ &= \frac{1/\omega C}{R} = \frac{1}{\omega CR}\end{aligned}$$

which is the same as Eq. (8.37). This obviously follows from the pythagorean relation between V, V_R and V_C. The expression for the impedance also follows directly from this relation.

$$\begin{aligned}V_m^2 &= I_m^2 R^2 + (I_m/\omega C)^2 \\ &= I_m^2[R^2 + (1/\omega C)^2] = I_m^2 Z^2\end{aligned}$$

where $Z = \sqrt{R^2 + (1/\omega C)^2}$ i.e. Eq. (8.36).

We will see later that currents and voltages in other (and often more complicated) ac circuits can be found and understood using the phasor method.

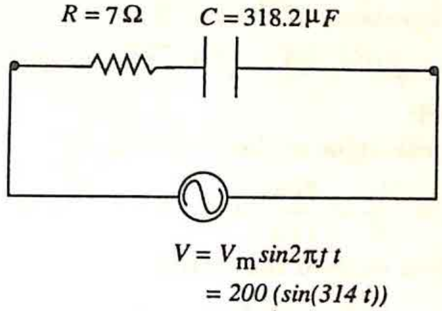
Example 8.3: The values of the circuit elements for an RC circuit and the voltage applied are as shown in figure (a) below. Calculate the following quantities: the capacitive reactance, the impedance, the maximum and rms currents, the phase angle, and voltages across various circuit elements. Describe the current and voltage graphically via the phasor diagram.

Answer:

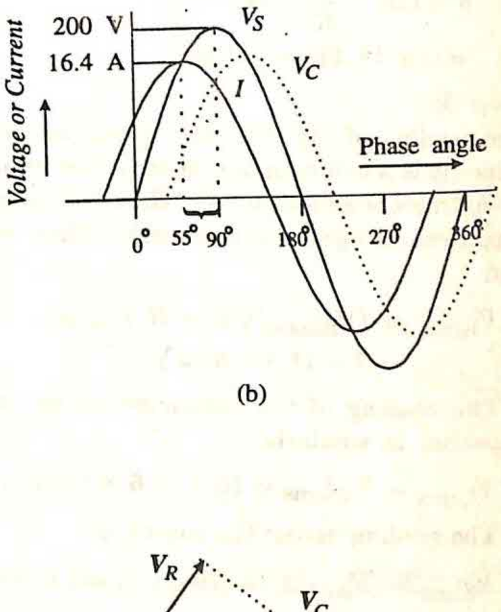
Step 1: Since the frequency of the supply is 50 Hz, its angular frequency $\omega = 2\pi\nu$ rad/s $= 2 \times 3.14 \times 50 = 314$ rad/s.

Step 2: The capacitive reactance

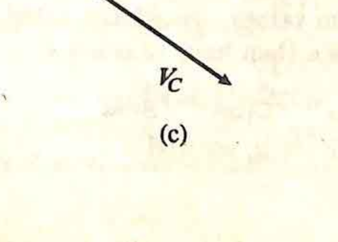
$$\begin{aligned}X_c &= \frac{1}{\omega C} = \frac{1}{314 \times 318.2 \times 10^{-6}}\Omega \\ &= \frac{10^6}{99962} \approx 10\Omega\end{aligned}$$



(a)



(b)



(c)

Step 3:
The impedance of the circuit is

$$Z = \sqrt{R^2 + X_c^2} = \sqrt{7^2 + 10^2} = 12.2\Omega$$

Step 4:
The peak value of the current is

$$I_m = \frac{V_m}{Z} = \frac{200}{12.2} = 16.4\text{A}$$

The rms value of the current is

$$I_{\text{rms}} = \frac{I_m}{\sqrt{2}} = \frac{16.4}{\sqrt{2}} = 11.6\text{A}.$$

Step 5:
The phase angle between the voltage and the current is

$$\theta = \tan^{-1}\frac{X_c}{R} = \tan^{-1}\frac{10}{7}$$
$$= \tan^{-1} 1.428 = 55^\circ \text{ lead}.$$

Step 6:
The reading of the voltmeter gives the rms value (it is a common and general feature of ac voltmeters and ammeters that they measure rms voltages and currents). Thus, we find

$$V_{\text{Rrms}} = (V_{\text{Rmax}}/\sqrt{2}) = R \times I_{\text{rms}}$$
$$= 7 \times 11.6 = 81.2\,\text{V}$$

The reading of the voltmeter across the capacitor is, similarly

$$V_{\text{Crms}} = X_C I_{\text{rms}} = 10 \times 11.6 = 116\text{V}.$$

The reading across the source is

$$V_{\text{Srms}} = (V_m/\sqrt{2}) = (200/\sqrt{2}) = 141.5\text{V}.$$

We check the consistency of these results. We have seen earlier that the *maximum* voltages across the resistor, the capacitor and the source satisfy the relation Eq. (8.8). The rms values, which are $(1/\sqrt{2})$ times the maximum values, should also satisfy this relation. We then have to check if

$$V_{\text{Rrms}}^2 + V_{\text{Crms}}^2 = V_{\text{Srms}}^2$$

Substituting, we find

$$\sqrt{(81.2)^2 + (116)^2} = (141.5) \text{ as expected!}$$

Let us now plot the voltage and the current waveforms (Fig. (b)) and their vector or phasor diagrams (Fig. (c)).

The applied voltage is $V = V_m \sin\omega t$ and the waveform V in Fig. (b) represents a sine wave.

The current leads the voltage by an angle $\theta = 55^\circ$ and is, therefore, represented as

$$I = I_m \sin(\omega t + 55^\circ).$$

This means that the current reaches its peak value $\theta = 55^\circ$ before the voltage reaches its maximum value. The charge Q and therefore V_C, the voltage across the capacitor *lags* behind the current by 90°.

Why should the current *lead* the voltage across the capacitor? The answer is obvious when you consider that the potential V_C is Q/C i.e. as charge Q is deposited on the capacitor plate due to the flow of current, the voltage V_C builds up *in step* with it.

The current flow *precedes* the voltage build up and when the current ceases to flow, the maximum charge has accumulated, giving maximum voltage. Now turn to Fig. (b) and you will see that the voltage is zero when the current is at its peak and reaches its maximum value when the current reaches its zero value.

Fig. (c) is the phasor representation of the voltages and currents. The applied voltage V is taken as the reference along the x-axis; all the other voltage vectors have been drawn to the same scale. As has been stated earlier, one may use either maximum or the rms values drawing these phasor diagrams.

8.6 Analytical study of the LR circuit

8.6.1 LR series circuit with a dc voltage

We now discuss what happens when a dc voltage is switched on to an LR circuit. The voltage equation is

$$L\frac{dI}{dt} + RI = V \tag{8.42a}$$

or

$$\frac{dI}{dt} + \frac{R}{L}I = \frac{V}{L} \tag{8.42b}$$

Solving this equation, we have

$$I = \frac{V}{R}\left[1 - e^{-(tR/L)}\right] \tag{8.42c}$$

Obviously the time constant here is $\tau = (L/R)$. Differentiating Eq. (8.42c), we have

$$\frac{dI}{dt} = \frac{V}{R}\left[-e^{-(tR/L)}\right]$$

The voltage across the inductor is, therefore

$$V_L = L\frac{dI}{dt} = L.\left(-\frac{R}{L}\right).\frac{V}{R}\left(-e^{-(tR/L)}\right)$$
$$= Ve^{-tR/L} = Ve^{-t/\tau} \tag{8.43}$$

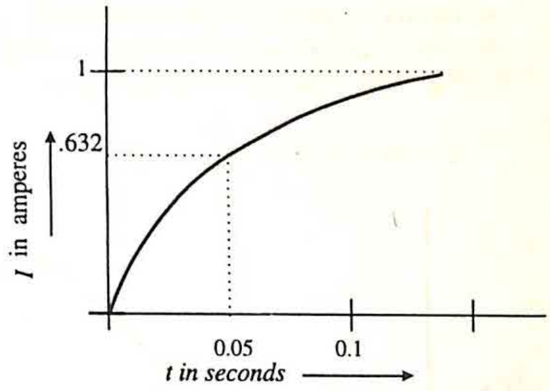
At $t = 0, V_L = V$ and at $t = \infty, V_L = 0$ and this corresponds to the experimental observation described in Section 8.3.

Example 8.4: A dc voltage of 10 V is switched on to a coil whose inductance is $L = 0.5$ H and which is in series with a resistance of 10Ω. Calculate the time constant, and plot the current and the voltage across the inductor.

Answer: The time constant $\tau = (L/R) = (0.5/10) = 0.05$s

$$I = \frac{V}{R}\left(1 - e^{-(t/\tau)}\right)$$
$$= \frac{10}{10}\left(1 - e^{-(t/0.05)}\right) = 1\left(1 - e^{-20t}\right) \text{ A.}$$

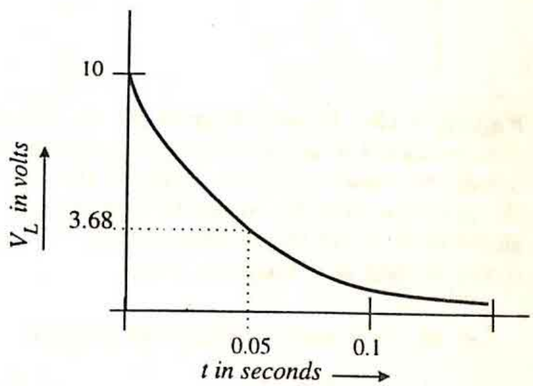
This is plotted in Fig. (a).



(a)

$$V_L = Ve^{-20t} = 10e^{-20t} \text{ V}$$

This is plotted in Fig.(b).



(b)

Note that the final current is $(V/R) = 1$A which remains constant at this value. The current builds up slowly if the time constant τ is large. This happens when the inductance is large or the resistance is small.

8.6.2 LR series circuit with ac voltage

Let us now consider an LR series circuit with an alternating sinusoidal voltage applied to it (Fig. 8.9). The voltage equation is

$$V_m \sin\omega t = RI + L\frac{dI}{dt} \quad (8.44)$$

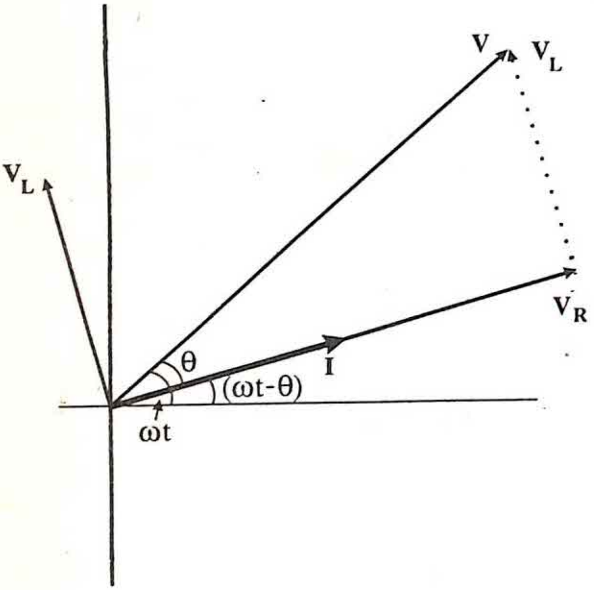


Figure 8.13: Phasor diagram for an LR circuit connected to an ac voltage. The voltage $\mathbf{V}_R$ across the resistor, and the current $\mathbf{I}$ through it, the potential drop $\mathbf{V}_L$ across the inductor ($\pi/2$) ahead of $\mathbf{V}_R$) and the resultant voltage $\mathbf{V}$ equal to the applied ac voltage are shown.

Let us once again assume the solution

$$I = I_m \sin(\omega t + \phi) \quad (8.45a)$$

so that

$$\frac{dI}{dt} = \omega I_m \cos(\omega t + \phi) \quad (8.45b)$$

Substituting Eq. (8.45a) and (8.45b) in Eq. (8.44), we have

$$\begin{aligned} V_m \sin\omega t &= RI_m \sin(\omega t + \phi) \\ &\quad + \omega L I_m \cos(\omega t + \phi). \end{aligned} \quad (8.46)$$

Usually ωL is represented as X_L and is described as *inductive reactance*. Multiplying and dividing by

$$Z = \sqrt{R^2 + X_L^2}, \text{ where } X_L = \omega L, \quad (8.47)$$

it follows that

$$\begin{aligned} V_m \sin\omega t &= Z I_m \left[\frac{R}{Z}\sin(\omega t + \phi) \right. \\ &\quad \left. + \frac{X_L}{Z}\cos(\omega t + \phi)\right] \end{aligned} \quad (8.48)$$

Just as in the case of RC circuit, we assume

$$\frac{R}{Z} = \cos\theta, \ \frac{X_L}{Z} = \sin\theta$$

and so

$$\tan\theta = \frac{X_L}{R} \quad (8.49)$$

Therefore

$$V_m \sin\omega t = Z I_m \sin(\omega t + \phi + \theta) \quad (8.50)$$

Once again, we have

$$V_m = Z I_m \text{ and } \phi + \theta = 0 \text{ or } \phi = -\theta. \quad (8.51)$$

This shows that the peak value of the current in a series LR circuit is

$$I_m = \frac{V_m}{Z} = \frac{V_m}{\sqrt{R^2 + (\omega L)^2}} \quad (8.52)$$

and the current *lags* the applied voltage by θ (note the minus sign) where

$$\theta = \tan^{-1}(X_L/R) = \tan^{-1}(\omega L/R). \quad (8.53)$$

We describe now the current and voltage in the LR circuit in terms of the phasor or vector diagram, just as we did for the RC circuit (Section 8.5, Fig. 8.12). Fig. 8.13 shows the voltage $\mathbf{V}_R$ across the resistance R at a particular time. Its magnitude is $I_m R$. The current $\mathbf{I}$ is in phase with this voltage, i.e. is directed the same way. The voltage drop V_L across the inductor L is $L(dI/dt)$ so that if $I = I_m \sin(\omega t + \phi)$, then $V_L =$

$(I_m\omega L)\cos(\omega t+\phi) = (I_m\omega L)\sin(\omega t+\phi+(\pi/2))$.

Thus the voltage $\mathbf{V}_L$ is 90° *ahead* of $\mathbf{V}_R$ or $\mathbf{I}$. This is shown in Fig. 8.13.

The magnitude of $\mathbf{V}_L$ is $(I_m\omega L)$. We now add $\mathbf{V}_L$ and $\mathbf{V}_R$ vectorially to give the total or applied ac voltage $\mathbf{V}$. It is also clear that the current $\mathbf{I}$ necessarily lags behind the voltage $\mathbf{V}$ by an angle $\theta = \tan^{-1}(\omega L/R)$ (e.g. Eq. (8.53)), and that the impedance $(V_m/I_m) = \sqrt{R^2+(\omega L)^2}$ (Eq. (8.52)).

Example 8.5: In the LR series circuit (Fig.a) where $R = 9\Omega, L = 20$ mH, $V = 200\sin\omega t$, and $\nu = 50$ Hz, find the root mean square values of the current in the circuit and the voltages across the circuit elements.

Answer: $X_L = \omega L = 2\pi\nu L = 314.1\times 20\times 10^{-3}\Omega = 6.28\Omega$ is the inductive reactance.

The impedance $Z = \sqrt{R^2+X_L^2} = \sqrt{9^2+6.28^2} = 10.97\Omega$

$I_m = (V_m/Z) = (200/10.97) = 18.23$A is the maximum current.

$\theta = \tan^{-1}(X_L/R) = \tan^{-1}(6.28/9) = 34.9°$ is the phase lag.

$$I_{\rm rms} = \frac{I_m}{\sqrt{2}} = \frac{18.23}{\sqrt{2}} = 12.89\text{A}$$

$V_{\rm Rrms} = RI_{\rm rms} = 9\times 12.89 = 116$ V is the voltage across the resistor.

$V_{\rm Lrms} = X_L I_{\rm rms} = 6.28\times 12.89 = 80.94$ V is the voltage across the inductor.

Check : The rms value of $V_{\rm Srms} = (200/\sqrt{2}) = 141.4$ V. Also

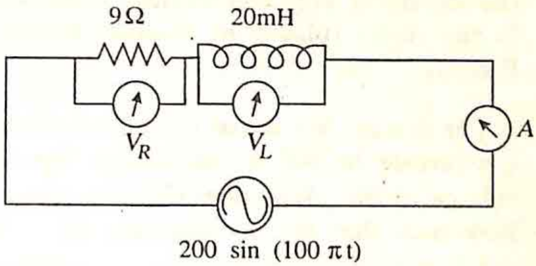
$$\sqrt{V_{\rm Rrms}^2 + V_{\rm Lrms}^2} = V_{\rm Srms}$$

or

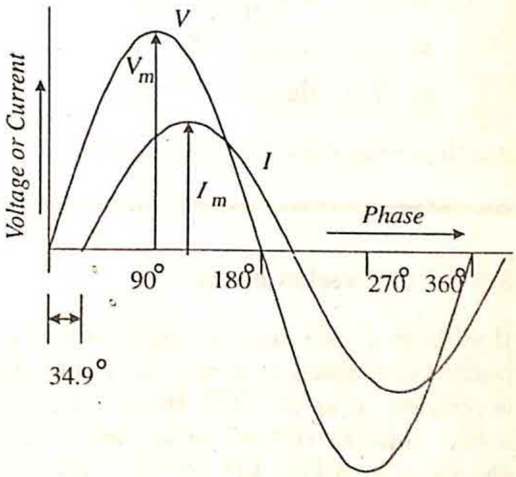
$$\sqrt{116^2+80.9^2}\text{ V} = 141.4\text{ V}$$

which confirms our earlier observations.

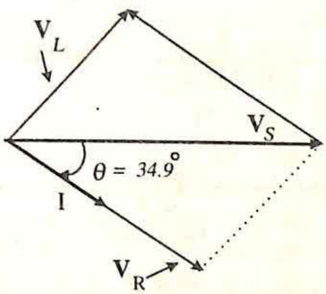
The voltages and currents are now represented as in Fig.(b). The current lags behind



(a)



(b)



(c)

the voltage by 34.9° which is also represented in the vector (phasor or Argand) diagram, Fig.(c).

The voltage $\mathbf{V}_L$ across the inductor *leads* the current by 90° or the current *lags* the voltage by 90°. Why does this happen? We have seen that the voltage induced in the coil $\mathcal{E} = -L(dI/dt)$. Since this opposes the applied voltage the voltage drop across the inductor is

$$\begin{aligned} V_L &= L\frac{\mathrm{d}I}{\mathrm{d}t} = L\frac{\mathrm{d}}{\mathrm{d}t}[I_m \sin(\omega t - \theta)] \\ &= \omega L I_m \cos(\omega t - \theta) \\ &= X_L I_m \sin(\omega t - \theta + \pi/2) \end{aligned}$$

which proves that $\mathbf{V}_L$ leads $\mathbf{I}$ by 90°.

8.7 LCR series circuit

If we have a resistance, an inductor and a capacitor connected in series, the combination is referred to as an *LCR* series circuit. Let a sinusoidal ac voltage be applied to such a circuit (Fig. 8.14). The voltage equation is

$$\begin{aligned} V_m \sin\omega t &= L\frac{\mathrm{d}I}{\mathrm{d}t} + RI + \frac{Q}{C} \\ &= L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C} \end{aligned} \qquad (8.54)$$

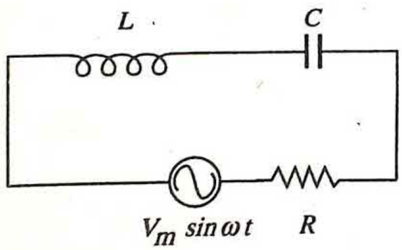


Figure 8.14: An *LCR* circuit, with all the elements, i.e. inductor, capacitor and resistor in series, connected to mains with voltage $V_m \sin\omega t$.

As was done on previous occasions, we assume the solution

$$Q = Q_m \sin(\omega t + \phi) \qquad (8.55a)$$

Differentiating, we have

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = Q_m\omega\cos(\omega t + \phi) \qquad (8.55b)$$

and

$$\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} = -Q_m\omega^2\sin(\omega t + \phi) \qquad (8.55c)$$

Substituting in Eq. (8.54) and rearranging the terms we have

$$V_m \sin\omega t = Q_m\omega\left[R\cos(\omega t + \phi) + \left(\frac{1}{\omega C} - \omega L\right)\sin(\omega t + \phi)\right] \qquad (8.56)$$

Here ωL is the inductive reactance designated as X_L, $(1/\omega C)$ is the capacitive reactance designated as X_C and Z is the impedance of the circuit defined as

$$Z = \sqrt{R^2 + (X_C - X_L)^2}. \qquad (8.57)$$

Multiplying and dividing Eq. (8.56) by Z, we have

$$V_m \sin\omega t = Q_m\omega Z\left[\frac{R}{Z}\cos(\omega t + \phi) + \frac{(X_C - X_L)}{Z}\sin(\omega t + \phi)\right] \qquad (8.58)$$

Again, we denote

$$\begin{aligned} (R/Z) &= \cos\theta \text{ and} \\ \frac{X_C - X_L}{Z} &= \sin\theta \\ \text{so that } \theta &= \tan^{-1}\frac{X_C - X_L}{R}. \end{aligned} \qquad (8.59)$$

Substituting in Eq. (8.58) we find

$$V_m \sin\omega t = I_m Z\cos(\omega t + \phi - \theta) \qquad (8.60a)$$

$$\text{so that} \quad V_m = I_m Z \qquad (8.60b)$$

$$\phi - \theta = -\frac{\pi}{2} \tag{8.60c}$$

or

$$\phi = -\frac{\pi}{2} + \theta. \tag{8.60d}$$

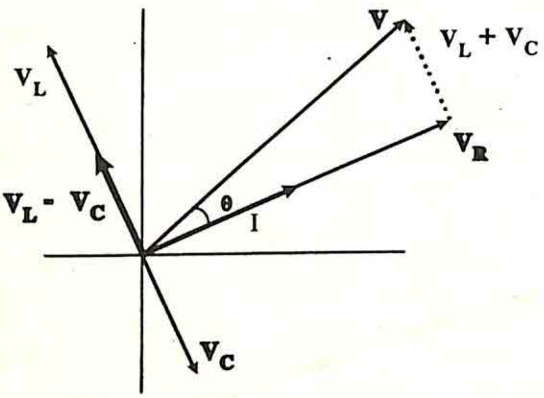


Figure 8.15: Vector or phasor diagram for series LCR circuit to which an ac voltage is applied. The voltage $\mathbf{V}_R$ across the resistor R and the electric current **I**, both point in the direction as shown. Perpendicular to this are $\mathbf{V}_L$ (leading by $(\pi/2)$) and $\mathbf{V}_C$ (lagging by $(\pi/2)$). The net voltage $(\mathbf{V}_L + \mathbf{V}_C)$ in the perpendicular direction (which has a magnitude $(V_L - V_C)$) and the final voltage $\mathbf{V} = \mathbf{V}_R + \mathbf{V}_L + \mathbf{V}_C$ are also shown.

The current

$$I = \frac{dQ}{dt} = Q_m \omega \cos(\omega t + \phi)$$
$$= I_m \sin(\omega t + \theta) \tag{8.61}$$

It is very important to note that the sign of θ will depend on whether $X_C > X_L$ or $X_L > X_C$ since

$$\theta = \tan^{-1}\{(X_C - X_L)/R\} \tag{8.62}$$

If $X_C > X_L, \theta$ is positive and

$$I = \frac{V_m}{Z} \sin(\omega t + \theta) \tag{8.63}$$

implies that the current *leads* the applied voltage.

If $X_L > X_C, \theta = \tan^{-1}\{(X_C - X_L)/R\}$ is negative and so

$$I = \frac{V_m}{Z} \sin[\omega t + (-\theta)]$$
$$= \frac{V_m}{Z} \sin(\omega t - \theta) \tag{8.64}$$

and the current *lags* the voltage.

Let us draw the vector diagram (Fig. 8.15) for the case when $X_L > X_C$. The voltage $\mathbf{V}_R$, as well as the voltages $\mathbf{V}_L$ and $\mathbf{V}_C$ which, respectively, lead and lag by 90°, are shown in the figure. The net voltage $\mathbf{V}_L - \mathbf{V}_C$ and the resultant $\mathbf{V}$($\equiv$ the applied ac voltage) are also shown there. We see that

$$|\mathbf{V}|^2 = R^2 I^2 + (X_C - X_L)^2 I^2 \tag{8.65a}$$

or

$$|\mathbf{V}| = \sqrt{R^2 + (X_L - X_C)^2}.I = Z.I \tag{8.65b}$$

The angle $\theta = \tan^{-1}\{(X_C - X_L)/R\}$ being negative, the current *lags* the voltage. This circuit is predominantly inductive. Since $X_L = 2\pi\nu L$ and $X_C = (1/2\pi\nu C)$, the reactances depend on the frequency of the supply ν. If L and C are kept constant, and we increase ν, X_L increases and X_C decreases so the same circuit may become inductive $(X_L > X_C)$ or capacitive $(X_L < X_C)$ depending on the value of the frequency of supply. There must be a frequency at which $X_L = X_C$. The inductive reactance exactly cancels the capacitive reactance. This is a very interesting situation which we shall examine in some detail. But before we do so, let us examine how a circuit absorbs, returns or dissipates power/energy. We shall find that whereas a resistor absorbs energy, an inductor and a capacitor connected to an ac supply do *not* absorb net energy over a cycle, though a current flows and there is a voltage drop. This result is quite surprising, and the

reason is discussed in the next section, where we look into power flow in such circuits.

8.8 Power flow in a circuit with ac voltage supply

8.8.1 A pure resistor across an ac voltage source

You have already learnt that the average power dissipated in a resistor is (see Eq. (8.7))

$$P = I_{\text{rms}}^2 R$$

where I_{rms} is the rms value of the current. When the voltage supply is sinusoidal, $I_{\text{rms}} = (I_m/\sqrt{2})$ and therefore

$$P = I_{\text{rms}}^2 R = (I_m^2/2)R.$$

If the circuit does not have any element other than the resistance R,

$$I_m = (V_m/R)$$

and also $I_{\text{rms}} = (V_{\text{rms}}/R)$ where V_{rms} is the rms voltage $= (V_m/\sqrt{2})$. It clearly follows that

$$P = (V_{\text{rms}}^2/R) = (V_m^2/2R) \tag{8.66}$$

in a circuit with a resistor and a sinusoidal voltage applied to it.

8.8.2 A pure capacitor across on ac voltage source

The net power/energy retained by a pure capacitor connected across an ac voltage source is zero. Whatever power a capacitor absorbs from the source in a quarter of a cycle, it returns in the next quarter. We show this in some detail now.

The instantaneous power P in a circuit with a capacitor only, and with an ac voltage applied to it is

$$P = (V_m \sin \omega t).(I_m \cos \omega t). \tag{8.67}$$

Since the current leads the voltage by $90°, I = I_m \cos \omega t$ if the applied voltage is $V_m \sin \omega t$. We thus have

$$\begin{aligned} P &= \frac{V_m I_m}{2} \sin 2\omega t, \\ &= V_{\text{rms}} I_{\text{rms}} \sin 2\omega t, \end{aligned} \tag{8.68}$$

(V_{rms} and I_{rms} are rms values). It is evident that $\sin 2\omega t$ is positive from $0-90°$ or during the first quarter cycle of the voltage/current waves. From $90°-180°$, i.e. the next quarter cycle, the power is negative. All the positive and negative power parts cancel each other as we shall see in Fig. 8.16. Eq. (8.68) can be obtained in another way. Since the energy E stored in a capacitor is $(Q^2/2C)$ (see Eqs (2.82) and (2.83)), the power P or rate of change of energy is

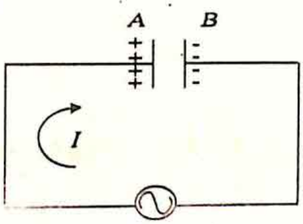
$$\begin{aligned} P &= \frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Q^2}{2C}\right) \\ &= \left(\frac{Q}{C}\right)\frac{\mathrm{d}Q}{\mathrm{d}t} = V_C I \end{aligned} \tag{8.69}$$

namely Eq. (8.68).

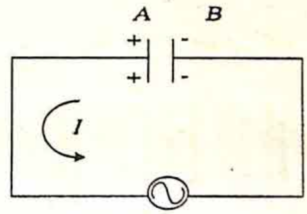
8.8.3 A pure inductor connected to an ac source

When a pure inductor is connected to an ac source, the net power absorbed in one complete cycle is zero.

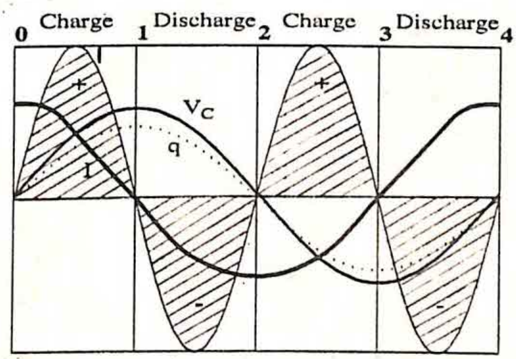
In an ideal inductor $R = 0$ and therefore θ or $\tan^{-1}(X_L/R)$, the angle by which the current through an inductor lags the voltage, is 90°. The expression for the current is $I_m \sin(\omega t - 90°) = -I_m \cos \omega t$. Just as in the case of a capacitor the power $P = -V_m I_m \sin \omega t \cos \omega t = -(V_m I_m/2) \sin 2\omega t = -VI \sin 2\omega t$ when integrated over a complete cycle becomes zero. This is illustrated in the Fig. 8.17.



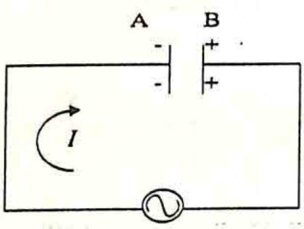
0-1 The current I flows as shown and from the maximum at 0 reaches a zero value at 1. The plate A is charged to positive polarity while negative charge q builds up in B reaching a maximum at 1 until the current becomes zero. The voltage $V_C = q/C$ is in phase with q and reaches a maximum value at 1. Current and voltage are both positive. So $P = V_C I$ is positive.
ENERGY IS ABSORBED FROM THE SOURCE FOR A QUARTER CYCLE AS THE CAPACITOR IS CHARGED.



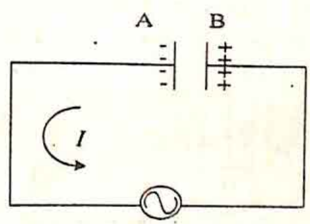
1-2 The current I reverses its direction. The accumulated charge is depleted i.e., the CAPACITOR IS DISCHARGED DURING THIS QUARTER CYCLE. The voltage gets reduced but is still positive. The current is negative. Their product, the power is negative.
THE ENERGY ABSORBED DURING THE 1/4 CYCLE **0-1** IS RETURNED DURING THIS QUARTER



One complete cycle of voltage/current Note that the current leads the voltage.

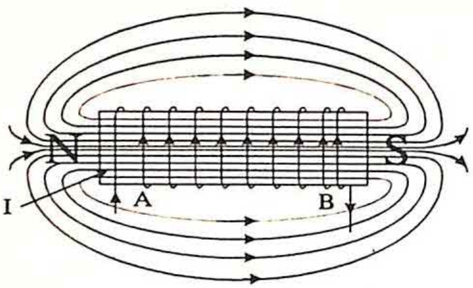


2-3 As I continues to flow from A to B, the Capacitor is CHARGED to reversed polarity i.e., the plate B acquires positive and A acquires negative charge. Both the current and the voltage are NEGATIVE. Their product P is POSITIVE. The capacitor ABSORBS ENERGY during this 1/4 cycle.

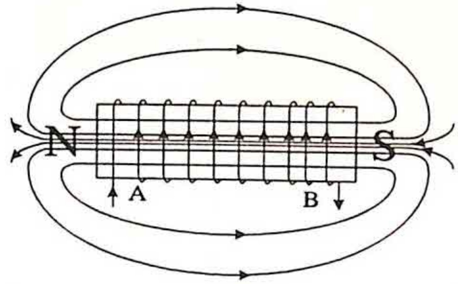


3-4 The current I reverses its direction at 3 and flows from B to A. The accumulated charge is depleted and the magnitude of the voltage V_C is reduced. V_C becomes zero at 4 when the capacitor is fully discharged. The power is negative and ENERGY ABSORBED DURING 2-3 IS RETURNED TO THE SOURCE. NET ENERGY ABSORBED IS ZERO.

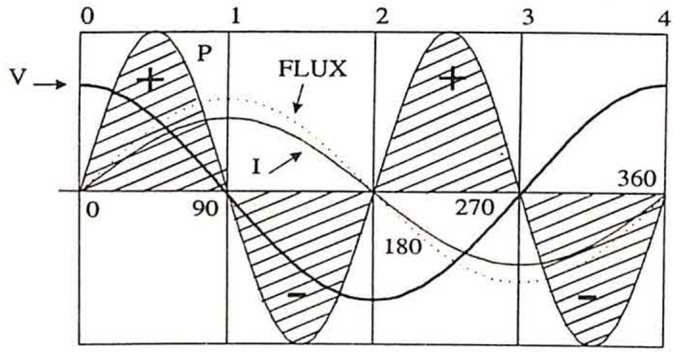
Figure 8.16: Charging and discharging of a capacitor



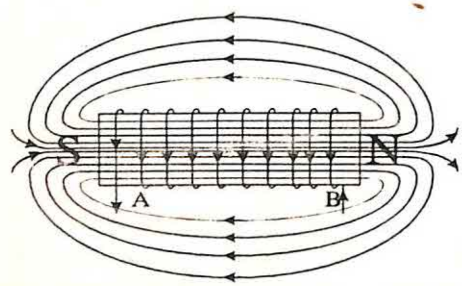
0-1 Current I through the coil entering at A increase from zero to a maximum value. Flux lines are set up i.e., the core gets magnetized. With the polarity shown voltage and current are both positive. So their product P is positive. ENERGY IS ABSORBED FROM THE SOURCE.



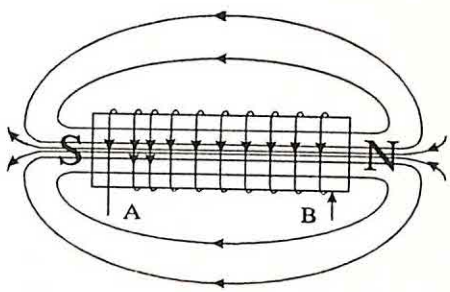
1-2 Current in the coil is still positive but is decreasing. The core gets demagnetized and the net flux becomes zero at the end of half cycle. The voltage V is negative (since $\mathrm{d}I/\mathrm{d}t$ is negative). The product of voltage and current is negative, and ENERGY IS BEING RETURNED TO SOURCE.



One complete cycle of voltage/current Note that the current lags the voltage.



2-3 Current I becomes negative i.e., it enters at B and comes out of A. Since the direction of current has changed, the polarity of the magnet changes. The current and voltage are both NEGATIVE. So their product P is POSITIVE. ENERGY IS ABSORBED.



3-4 Current I decreases and reaches its zero value at 4 When core is demagnetized and flux is zero, the voltage IS POSITIVE but the current is NEGATIVE. The power is therefore NEGATIVE, ENERGY ABSORBED DURING THE 1/4 CYCLE 2-3 IS RETURNED TO THE SOURCE.

Figure 8.17: Magnetization and demagnetization of an inductor

8.8.4 Power consumed in a LCR circuit: The Power Factor

We have seen in our earlier discussions (Figs. 8.16 and 8.17) that an inductor and a capacitor either consume or return power at an instant of time, but taken over a complete cycle the net power consumed averages to zero. Resistor, however, never returns any power back to the source (Fig. 8.5) and consumes an average power of $I_{\rm rms}^2 R$ (Eq. (8.7)).

Let us see what happens when an inductor, a capacitor and a resistor in series are connected to an ac voltage source. We have seen (Eq. (8.63)) that the current in an LCR circuit is $I = I_m \sin(\omega t + \theta)$ where $\theta = \tan^{-1}[(X_C - X_L)/R]$. The *instantaneous* power dissipation P is

$$\begin{aligned} P &= VI = (V_m \sin\omega t) \times (I_m \sin(\omega t + \theta)) \\ &= \frac{V_m I_m}{2}[\cos\theta - \cos(2\omega t + \theta)]. \end{aligned} \tag{8.70}$$

Over a cycle, (actually over a half cycle) the second term in the square brackets of Eq. (8.70b) averages to zero, the positve half of the cosine cancelling the negative half.

The average power loss in an LCR circuit is, therefore

$$\begin{aligned} P_{av} &= \frac{V_m I_m}{2}\cos\theta \\ &= \frac{V_m}{\sqrt{2}}\frac{I_m}{\sqrt{2}}\cos\theta \\ &= V_{\rm rms} I_{\rm rms}\cos\theta \end{aligned} \tag{8.71}$$

We notice the average power dissipated depends not only on the voltage and current, but also on the cosine of the phase angle θ between them.

The quantity $\cos\theta$ is called the *power factor*. We can directly show that the power P_{av} is consumed by the resistor only. The instantaneous power consumed by the resistor is

$$\begin{aligned} V_R I &= I^2 R = I_m^2 R \sin^2(\omega t + \theta) \\ &= (I_m R) I_m \sin^2(\omega t + \theta) \\ &= (V_m I_m \cos\theta)\sin^2(\omega t + \theta) \\ &= \frac{V_m I_m}{2}\cos\theta(1 - \cos 2(\omega t + \theta)) \end{aligned} \tag{8.72}$$

(since $V_m \cos\theta = V_R = I_m R$). Comparing Eq. (8.70) and (8.72) we can see that the instantaneous power in the entire LCR circuit is different from the instantaneous power in the resistor. But average values of these expression taken over one cycle is the same. This clearly shows that in a LCR circuit the average power is only consumed in the resistor and is equal to

$$\begin{aligned} P_{av} &= \frac{V_m}{\sqrt{2}}.\frac{I_m}{\sqrt{2}}\cos\theta \\ &= V_{\rm rms} I_{\rm rms}\cos\theta \quad (8.73a) \\ &= \frac{V_{\rm rms}^2}{Z}\cos\theta \quad (8.73b) \end{aligned}$$

It follows from the phasor diagram (see Fig. 8.15) that $VI\cos\theta = (V\cos\theta)I = V_R I = I^2 R$ (if we assume that V and I represent rms rather than the peak values) and proves that power loss takes place in the resistor.

One can make the general statement that power loss in an ac circuit (or *active power* as it is occasionally called by engineers) is the product of the rms value of the current multiplied by the *component in phase* with the rms voltage.

8.9 Resonance and oscillations

8.9.1 Resonance

We have earlier mentioned that in a LCR circuit (Fig. 8.14), depending on the values of L and C and the frequency of supply ν, the circuit may be inductive ($X_L > X_C$) or capactive ($X_L < X_C$).

If we change the frequency of supply, (or the values of inductor or capacitor) there may be a situation when

$$X_L = X_C$$
$$\text{or} \quad \omega L = \frac{1}{\omega C}$$
$$\text{or} \quad \omega^2 = \frac{1}{LC}$$

The angular frequency at which this happens is given a symbol ω_0. We have

$$\omega_0 = 2\pi\nu_0 = \frac{1}{\sqrt{LC}} \quad (8.74a)$$

$$\text{therefore,} \ \nu_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC}} \quad (8.74b)$$

If the circuit is neither inductive nor capacitative, it is only resistive. This is called the resonant condition and ν_0 is the resonance frequency or the natural frequency of the circuit.

At resonance the impedance $Z = \sqrt{R^2 + (X_C - X_L)^2} = R$ since $X_L = X_C$. It is clear that the impedance is a minimum when $X_C = X_L$. For any other case, i.e $X_L > X_C$ or $X_L < X_C$, $(X_C - X_L)^2$ is a positive quantity which adds to R^2 and increases the value of Z.

Since Z is a minimum, the current $I_{\text{rms}} = (V_{\text{rms}}/Z) = (V_{\text{rms}}/R)$ is maximum at resonance.

The power factor angle θ in a LCR circuit is

$$\theta = \tan^{-1} \frac{(X_C - X_L)}{R}$$

Since $(X_C - X_L) = 0, \theta = \tan^{-1} 0$ is equal to zero. i.e the phase angle between the applied voltage and current is zero. We therefore find that *at resonance the voltage and current are in phase.*

Since $\cos\theta = \cos 0^\circ = 1$, the power factor is unity.

The phase angle is zero at resonance, therefore, the power loss P is also maximum.

The phasor diagram Fig.(8.15) illustrates the above features. At resonance, the vector $\mathbf{V}_C$ and $\mathbf{V}_L$ are opposite (as always) and equal. This is a condition of minimum impedance, since any vectorial addition to $\mathbf{V}_R$ will result in a larger length $\mathbf{V}$ and hence large Z.

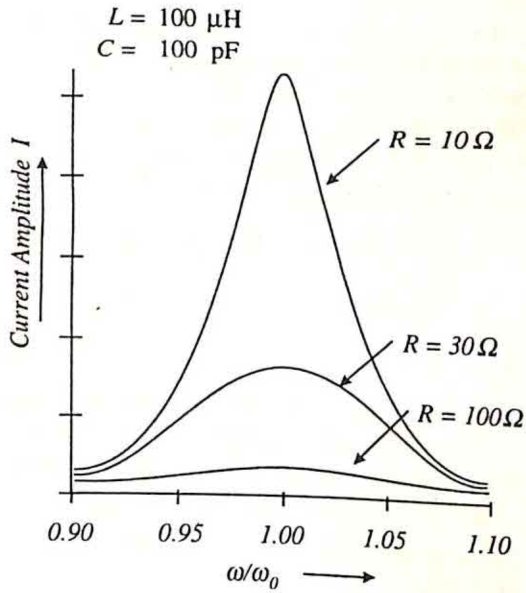


Figure 8.18: Resonance curves for forced oscillations in an LCR circuit like one shown in Fig.(8.14). The maximum current I_m is shown as a function of frequency ω in units of ω_0. For the values of $L = 100\mu$H and $C = 100p$F indicated, $\omega_0 = \sqrt{10} \times 10^5$ rad/s. The three curves are for resistances $R = 10\Omega$, 30 Ω and 100 Ω, respectfvely. The broadening of the resonance peak is clearly seen as R increases, though ω_0 remains unchanged.

We have, therefore, learnt that

(a) Resonance occurs in a LCR circuit

when
$|X_L - X_C| = 0$ or $\omega = (1/\sqrt{LC})$.

(b) The current reaches a maximum value of (V/R) at resonant condition.

(c) The power dissipated in the circuit is maximum and is equal to V_{rms}^2/R.

(d) The current is in phase with the voltage or the power factor is unity ($\cos\theta = 1$ when $\theta = 0$) (see Fig. 8.15).

We notice in Fig. 8.18 that the current/frequency curve becomes very sharp if the resistance of the circuit is very low. At resonant frequency, the current shoots up to a large value. If the frequency is a little less than ν_0 or a little more, the current becomes significantly less. This is not the case when the resistance of the circuit is large. The current/frequency curve is flat and the current does not change very much if the frequency deviates from ν_0. We can state this in a different manner. We can say that *the LCR circuit is more selective when resistance of the circuit is low.*

We may understand the selective property of a circuit better when we consider an everyday activity like tuning a radio set. By turning a knob what we are trying to do is to make the resonant or natural frequency of a circuit coincide with the frequency transmitted by the antenna of a particular broadcasting station. In fact we are seeking resonance. In a metropolitan area there are many signals in the air whose frequencies are close to each other so that to tune into one particular frequency the circuit should be highly selective.

The selectivity or sharpness of a resonant circuit is measured by what is called the Quality factor Q_0 of a coil which is given by

$$Q_0 = (\omega_0 L/R) \tag{8.75}$$

If R is low or L is large the quality factor is large implying that the circuit is more selective. Figure 8.18 shows the current – frequency curve for a typical LCR circuit. We see that while the current is always a maximum at $\omega = \omega_0$, the sharpness of this maximum decreases with increasing R.

8.9.2 LC Oscillations

Let us now examine the interesting circuit condition in which the resistance in the LCR circuit is zero. This is an ideal condition when we have an ideal inductor which has zero resistance. Devices can, however, be made which almost completely cancel the resistance of a circuit, to produce what is called an oscillator. You will study this later in this book.

What happens in a LC circuit having no resistance may be visualized with a mechanical analogue.

A block of mass m is connected through an elastic spring to a rigid wall and placed on a frictionless surface. If the block is pulled and let go, it will oscillate back and forth and if there is no friction, the block will go on oscillating forever (Newton's first law of motion). It is interesting to study how the mass and the spring exchange energy between them. When the mass moves to the extreme left position and stops, its kinetic energy $(1/2)mv^2$ is zero. Energy is stored as potential energy $(1/2)kx^2$ where k is the spring constant. As the spring pulls the mass back, it loses its potential energy and the mass gains kinetic energy which is maximum at the centre (where the spring is not stretched and its potential energy is zero). The inertia now makes the block move to the right when the spring is compressed and pushes the mass away. The mass slows down, losing its kinetic energy while the spring stores potential energy on being compressed. At any instant the sum

of the kinetic and potential energy remains a constant since it is not dissipated due to frictional resistance.

The equation of motion of the mass is given as

$$m\frac{d^2x}{dt^2} + kx = 0 \text{ or } \frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \quad (8.76)$$

The mass executes simple harmonic motion and the displacement x will have a solution of

$$x = A\sin\omega t$$

from which we can see, on differentiating both sides, that

$$\frac{dx}{dt} = A\omega \cos\omega t.$$

Differentiating again

$$\frac{d^2x}{dt^2} = -A\omega^2 \sin\omega t = -\omega^2 x$$

since $x = A\sin\omega t$. We now have

$$\frac{d^2x}{dt^2} + \omega^2 x = 0.$$

Comparing with Eq. (8.76) it follows that

$$\omega^2 = \frac{k}{m} \text{ or } \omega = \sqrt{\frac{k}{m}} \quad (8.77a)$$

$$\text{Therefore } \nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}}. \quad (8.77b)$$

i.e. the frequency at which the mass spring oscillates and oscillations continue *undamped*.

In an LC circuit something similar happens. Unlike previous cases we do not apply any voltage. In this case, the capacitor is initially charged (like we pulled the mass and let it go) and the current oscillates back and forth in the LC circuit. The equation for this is

$$L\frac{dI}{dt} + \frac{Q}{C} = 0$$

$$\text{or } L\frac{d^2Q}{dt^2} + \frac{Q}{C} = 0 \left(\text{since } I = \frac{dQ}{dt}\right)$$

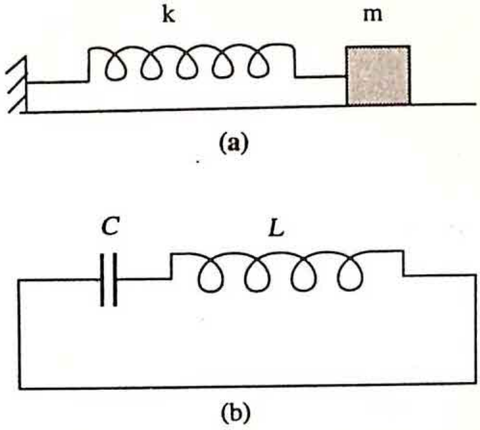


Figure 8.19: Mechanical oscillator similar to an oscillating LC circuit. A harmonic oscillator with a spring k stretched by a mass m (Fig.a) is similar to a capacitor C in series with an inductor L (Fig.b). The stretched spring is like the charged capacitor, and the moving mass is like the inductor through which a changing current flows.

$$\text{or } \frac{d^2Q}{dt^2} + \frac{1}{LC}\cdot Q = 0. \quad (8.78)$$

Note the similarity between this equation and the one for the mass and spring (Eq. (8.76)). The frequency of oscillation is given by

$$\omega = \sqrt{\frac{1}{LC}} \text{ or } \nu = \frac{1}{2\pi}\sqrt{\frac{1}{LC}} \quad (8.79)$$

which is nothing but the *natural frequency or resonant frequency of the circuit.*

Fig. 8.19 shows electrical and mechanical oscillations.

When the capacitor has a charge there is a stored energy of (1/2) (Q^2/C) in the electric field in the capacitor. This is like the elastic energy in the spring. The capacitor begins to discharge and a current I flows through the inductor, energy builds up in the magnetic field around the inductor, the magnetic energy is $(1/2)LI^2$ which is analogous to kinetic energy in the case of mass-spring. As

the current flows, charge builds up on the capacitor and the electric field builds up as the magnetic field declines. This exchange of energy keeps the current oscillating.

Example 8.6: A sinusoidal voltage of peak value 283 V and frequency 50Hz is applied to a series LCR circuit in which $R = 3\Omega, L = 25.48$ mH, $C = 7.96 \times 10^{-4}$ F. Find

(a) The rms voltage

(b) X_L and X_C

(c) the impedance Z

(d) the peak current and the phase angle

(e) the rms value of the current and voltage across the circuit elements

(f) the power dissipated in the circuit

(g) the power factor

(h) the power input

(i) the frequency of supply at which resonance occurs

(j) the impedance at resonant condition

(k) the current at resonant condition and

(l) the power dissipated at resonant condition

Answer:

(a) $V_{\rm rms} = (V_m/\sqrt{2})$ for ac voltage. Therefore,

$$V_{\rm rms} = (283/1.414) \simeq 200\,V$$

(b) $X_L = \omega L = 2\pi\nu L$

$$= 2 \times 3.141 \times 50 \times 25.48 \times 10^{-3}$$

$$= 8\Omega$$

$$X_c = \frac{1}{\omega C} = \frac{1}{2\pi\nu C} = \frac{1}{2 \times 3.141 \times 50} \times \frac{1}{\times 7.96 \times 10^{-6}} = 4\Omega$$

(c) $Z = \sqrt{R^2 + (X_L - X_C)^2}$

$$= \sqrt{3^2 + (8-4)^2} = \sqrt{9+16} = 5\Omega$$

(d) $I_m = (283/5) = 56.6$A,

$$\theta = \tan^{-1}[(X_L - X_C)/R] = \tan^{-1}(4/3) = 53.13^\circ$$

(e) $I_{\rm rms} = I_m/\sqrt{2} = 56.6/1.414 = 40$ A,

The current *lags* the voltage since $X_L > X_C$.

$$V_{\rm Crms} = X_C I_{\rm rms} = 4 \times 40 = 160{\rm V}$$
$$V_{\rm Lrms} = X_L I_{\rm rms} = 8 \times 40 = 320{\rm V}$$
$$V_{\rm Rrms} = R I_{\rm rms} = 3 \times 40 = 120{\rm V}$$

We check the consistency of these results

$$\sqrt{(V_{\rm Rrms})^2 + (V_{\rm Lrms} - V_{\rm Crms})^2} = \sqrt{(120)^2 + (320-160)^2} = 200{\rm V} = V_{\rm rms}.$$

(f) $P = (V_{\rm rms}/Z)^2 R = I_{\rm rms}^2 R = 40^2 \times 3 =$ 4800 W.

(g) Power factor $= \cos\theta = \cos 53.13^\circ =$ 0.6

(h) The power input $= V_{\rm rms} I_{\rm rms} \cos\theta = 200 \times 40 \times 0.6 = 4800$ W (Note that the power input is the same as the power dissipated in the resistor since the capacitor and the inductor do not absorb or produce any net power).

(i) For resonance,

$$\begin{aligned}\omega_0 &= \frac{1}{\sqrt{LC}}\\ &= \frac{1}{\sqrt{25.48\times10^{-3}\times7.96\times10^{-4}}}\\ &= \frac{10^4}{\sqrt{2028}} = 222.2\text{rad/s}\\ \omega_0 &= 2\pi\nu_0.\end{aligned}$$

Therefore

$$\nu_0 = \frac{222.2}{2\times3.141} = 35.4\,\text{Hz}$$

(j) The impedance Z at resonant condition is equal to the resistance

$$Z = R = 3\Omega$$

(k) The rms current at resonance

$$\begin{aligned}&= \frac{V_{\text{rms}}}{Z} = \frac{V_{\text{rms}}}{R}\\ &= (200/3) = 66.67A\end{aligned}$$

(l) The power consumed at resonance

$$\begin{aligned}P &= I_{\text{rms}}^2\times R\\ &= 66.67^2\times3 = 13.33\text{kW}\end{aligned}$$

At resonance

$$\begin{aligned}P &= \frac{V_{\text{rms}}^2}{Z^2}\times R\\ &= \frac{200^2}{3} = 13.33\,\text{kW}\end{aligned}$$

Summary

1. For an alternating current $I = I_m \sin \omega t$ passing through a resistor R, the average power loss P (averaged over a cycle) due to joule heating is $(1/2)I_m^2R$. This is often expressed in terms of the root mean squared (rms) current $I_{\text{rms}} = (I_m/\sqrt{2})$ as $P = (1/2)I_{\text{rms}}^2 R$. The rms voltage for this circuit is defined as $V_{\text{rms}} = (V_m/\sqrt{2}) = (I_m R/\sqrt{2})$.

2. In A.C. circuits, in addition to resistors, two elements: inductors and capacitors are relevant. The voltage-current relation for an inductor of self-inductance L is

   $$V = L\frac{di}{dt}$$

   The voltage-current relation for a capacitor of capacitance C is

   $$I = C\frac{dV}{dt}$$

   These relations should be contrasted with that for a resistor;

   $$V = I\ R$$

3. The magnetic flux Φ linking a coil due to the current I in the coil itself can be written as $\Phi = LI$ where the constant L is called the self inductance of the coil. Any change in I causes a change in Φ inducing an emf opposing the change ('back emf'). For a long solenoid of length ℓ, area of cross section A, and number of turns per unit length n, the self-inductance is given by

   $$L = \mu_0 n^2 \ell A$$

   Work done by external sources in building up current in an inductor L from 0 to I is (1/2) LI^2.

4. The magnetic flux Φ_2 through coil 2 due to current I_1 in coil 1 may be written as,

   $$\Phi_2 = M_{12} I_1$$

   where M_{21} is a constant that depends on the geometry of the coils: their size, shape, relative distance and orientation. The induced emf $\mathcal{E}_1$ in the coil 2 due to change in I_1 is given by Faraday's law:

   $$\mathcal{E}_2 = -\frac{d\Phi_2}{dt} = -M_{12}\frac{dI_1}{dt}$$

   M_{21} is similarly defined. The relation $M_{12} = M_{21} (= M)$ is true in general. M is called the mutual inductance between the two coils.

5. For a capacitor C connected in series with a resistor R across a DC source, the time constant τ is

$$\tau = C\ R$$

τ is a measure of the rate at which the charge of a capacitor grows from zero to its full value during charging; it also determines the rate at which the capacitor gets discharged.

For an inductor L connected in series with a resistor R across a DC source, the time constant τ is

$$\tau = \frac{L}{R}$$

τ is a measure of the rate at which the current in the circuit rises from zero to its final value when the circuit is switched on, or a measure of the rate at which the current drops to zero when the circuit is switched off.

6. For a capacitor C across an ac source: $V(t) = V_0 \sin \omega t$, the current I is

$$I = \omega C V_0 \sin\left(\omega t + \frac{\pi}{2}\right)$$

i.e the current leads the voltage by a phase angle of $\pi/2$ and the effective resistance of C is $1/\omega C$.

For an inductor L across an ac source whose voltage varies as $V(t) = V_0 \sin \omega t$, the current I is:

$$I = \frac{V_0}{\omega L} \sin\left(\omega t - \frac{\pi}{2}\right)$$

i.e the current lags behind the voltage by a phase angle $\pi/2$ and the effective resistance of L is ωL.

For a resistor R, current and voltage are always in phase.

7. For a series LCR circuit across an ac source $V(t) = V_0 \sin \omega t$, the current I is given by

$$I = I_0\ \sin(\omega t - \delta) = \frac{V_0}{\left\{R^2 + (\omega L - (1/\omega C))^2\right\}^{1/2}}$$

where

$$\tan \delta = \frac{\omega L - (1/\omega C)}{R}$$

The circuit exhibits resonance i.e the amplitude of the current is a peak at the resonant frequency

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

The Q-factor defined by

$$\frac{1}{Q_0} = \frac{R}{\omega_0 L}$$

is an indicator of the sharpness of the resonance, the higher value of Q_0 indicating sharper peak in the current.

8. When an ac source $V(t) = V_0 \sin \omega t$ is connected to an inductor L or a capacitor C, no net power is consumed in a *complete* cycle, since the current and voltage have a phase difference of magnitude $\pi/2$. When the source is connected to a resistor R, average power consumed in a full cycle is

$$P = \frac{V_0^2}{2R} = V_{\text{rms}} I_{\text{rms}}$$

where rms stand for 'root mean squared':

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}; \; I_{\text{rms}} = \frac{I_0}{\sqrt{2}}; \; I_0 = \frac{V_0}{R}$$

9. A circuit containing inductor L and capacitor C (initially charged) with no ac source, and no resistor exhibits free oscilaltions. The charge $Q(t)$ of the capacitor satisfies the equation of simple harmonic motion:

$$\frac{d^2Q(t)}{dt^2} + \frac{1}{LC} Q(t) = 0$$

so that the frequency ω of free oscillations is: $\omega = (1/\sqrt{LC})$. The *sum* of the energy stored in the inductor and the capacitor is constant in time.

Exercises

8.1 A 100 Ω resistor is connected to a 220 V, 50 Hz ac supply.

(a) What is the rms value of current in the circuit?

(b) What is the net power consumed over a full cycle?

8.2 (a) The peak voltage of an ac supply is 300 V. What is its rms voltage?

(b) The rms value of current in an ac circuit is 10 A.

What is the peak current?

Note: When the voltage of an ac supply is specified, it is understood to refer to its rms voltage.

8.3 What is the self-inductance of a solenoid of length 40 cm, area of cross-section 20 cm^2 and total number of turns 800?

8.4 A solenoid of length 50 cm with 20 turns per cm and area of cross-section 40 cm^2 completely surrounds another co-axial solenoid of the same length, area of cross-section 25 cm^2 with 25 turns per cm. Calculate the mutual inductance of the system.

8.5 A charged capacitor of capacitance 40 μF is discharged through a 50 Ω resistor. Determine the time- constant of the circuit and explain what it signifies.

8.6 A circuit containing a 30 mH inductor in series with a 60 Ω resistor is connected to a dc supply. Determine the time-constant of the circuit and explain what it signifies.

8.7 A 44 mH inductor is connected to 220 V, 50 Hz ac supply. Determine the rms value of the current in the circuit.

8.8 A 60 μF capacitor is connected to a 110 V, 60 Hz ac supply. Determine the rms value of the current in the circuit.

8.9 In 8.7 and 8.8, what is the net power absorbed by each circuit over a complete cycle. Explain your answer.

8.10 Show that a series LCR circuit driven by an ac source exhibits resonance at $\omega_r = 1/\sqrt{LC}$.

8.11 Obtain the resonant frequency ω_r of a series LCR circuit with $L = 2.0$ H, $C = 32\mu$F and $R = 10\Omega$.

What is the Q-value of this circuit?

8.12 Why is a choke coil needed in the use of fluroscent tubes with AC mains? Why can we not use an ordinary resistor instead of the choke coil?

8.13 Show that the angular frequency of free oscillations of an LC circuit is equal to $1/\sqrt{LC}$.

8.14 Show that in the free oscillations of an LC circuit, the sum of energies stored in the capacitor and inductor is constant in time.

8.15 A charged 30 μF capacitor is connected to a 27 mH inductor. What is the angular frequency of free oscillations of the circuit?

8.16 Suppose the initial charge on the capacitor in 8.15 is 6 mC. What is the total energy stored in the circuit initially? What is the total energy at later times?

8.17 A series LCR circuit with $R = 20\Omega, L = 1.5$H and $C = 35\mu$F is connected to a variable-frequency 200 V ac supply. When the frequency of the supply equals the natural frequency of the circuit, what is the average power transferred to the circuit in one complete cycle?

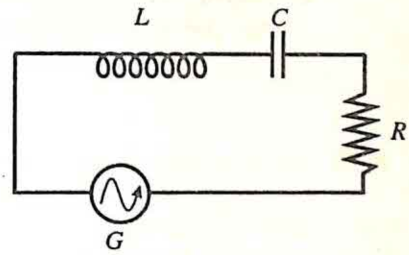
8.18 A radio can tune over the frequency range of a portion of MW broadcast band: (800 kHz to 1200 kHz) If its LC circuit has an effective inductance of 200 μH, what must be the range of its variable condenser?

Hint: For tuning, the natural frequency i.e. the frequency of free oscillations of the LC circuit should equal the frequency of the radiowave].

8.19 Figure shows a series LCR circuit connected to a variable frequency 230 V source. $L = 5.0$ H, $C = 80\mu$F, $R = 40\Omega$.

(a) Determine the source frequency which drives the circuit in resonance.

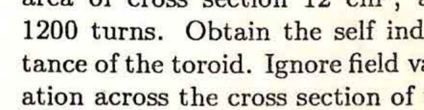
(b) Obtain the impedance of the circuit and the amplitude of current at the resonating frequency.



(c) Determine the rms potential drops across the three elements of the circuit. Show that the potential drop across the LC combination is zero at the resonating frequency.

Additional Exercises

8.20 Figure shows a short solenoid of length 4 cm, radius 2.0 cm and number of turns 100 lying inside on the axis of a long solenoid, 80 cm length and number of turns 1500. What is the flux through the long solenoid if a current of 3.0A flows through the short solenoid? Also obtain the mutual inductance of the two solenoids.



8.21 (a) A toroidal solenoid with an air core has an average radius of 15 cm, area of cross section 12 cm^2, and 1200 turns. Obtain the self inductance of the toroid. Ignore field variation across the cross section of the toroid.

(b) A second coil of 300 turns is

wound closely on the toroid above. If the current in the primary coil is increased from zero to 2.0A in 0.05s, obtain the induced emf in the second coil.

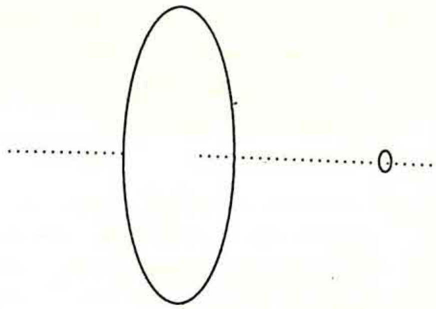
8.22 (a) A circuit contains two inductors in series, with self inductances L_1 and L_2 and mutual inductance M. Obtain a formula for the equivalent inductance in the circuit.

(b) Two inductors of self inductances L_1 and L_2 are connected in parallel. The inductors are so far apart that their mutual inductance is negligible. What is the equivalent inductance of the combination?

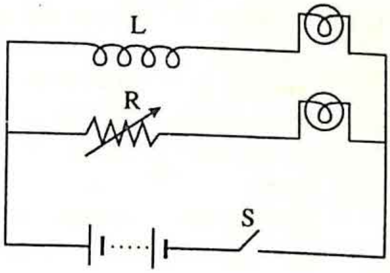
8.23 (a) Show that the energy stored in an inductor i.e. the energy required to build current in the circuit from zero to I is $(1/2)LI^2$ where L is the self inductance of the circuit.

(b) Extend this result to a pair of coils of self inductances L_1 and L_2, and mutal inductance M. Hence obtain the inequality $M^2 < L_1 L_2$.

8.24 A circular loop of radius 0.3 cm lies parallel to a much bigger circular loop of radius 20 cm. The centre of the small loop is on the axis of the bigger loop. The distance between their centres is 15 cm. (a) What is the flux linking the bigger loop if a current of 2.0A flows through the smaller loop? (b) Obtain the mutual inductances of the two loops.



8.25 Answer the following questions:



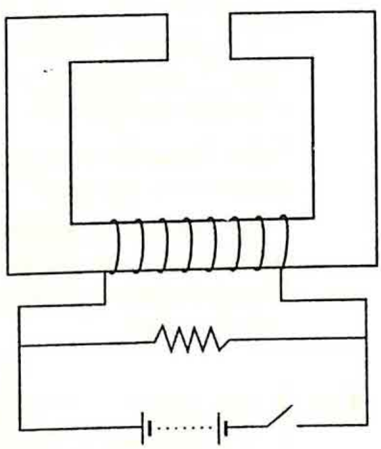
(a) Figure above shows an inductor L and a resistance R connected in parallel to a battery through a switch. The resistance of R is the same as that of the coil that makes L. Two identical bulb are put in each arm of the circuit. (i) Which of the bulbs lights up earlier when S is closed? (ii) Will the bulbs be equally bright after some time?

(b) An iron bar falling vertically through the hollow region of a thick cylindrical shell made of copper experiences retarding force. What can you conclude about the iron bar?

(c) A coil is wound on an iron core and looped back on itself so that the core has two sets of closely wound wires in series

carrying current in the opposite senses. What do you expect about its self inductance? Will it be large or small?

(d) An electron moves in a circle with uniform speed in a stationary magnetic field normal to the plane of the circle. If the field magnitude is made to increase with time, will the electron speed up or speed down? Will it continue to revolve in the same circle?



(e) A small resistor (say a lamp) is usually put in parallel to the current carrying coil of an electromagnet (See figure above). What purpose does it serve?

8.26 Two capacitors, 4 μF and 6 μF in series, are connected through a resistance of 10 kΩ to a 18 V battery of negligible internal resistance. After a time of about 10 s, the battery is disconnected and capacitors are allowed to discharge through the resistance. Determine the voltage across each capacitor after a time lapse of 48 ms.

8.27 Two circuit (A) and (B) connected to identical dc sources (emf = 10 V) differ greately in their self inductance. Circuit A has large self inductance = 10 H, while the self inductance of B is much smaller, equal to 0.05H. The total external resistance in each circuit (which includes the resistance of the inductor itself) is 40Ω.

(a) Are the steady current values in each circuit equal? If so what is the value?

(b) Compare the times required for the currents in the two circuits to reach [1−(1/e)] of their steady value?

(c) Which circuit requires greater energy consumption of the source to build up its current to the steady value?

(d) After the steady state is reached, do the circuits dissipate the same power in the form of heat?

8.28 An LC circuit contains a 20 mH inductor and a 50 μF capacitor with an initial charge of 10mC. The resistance of the circuit is negligible. Let the instant the circuit is closed be $t = 0$.

(a) What is the total energy stored initially? Is it conserved during

LC oscillations?

(b) What is the natural frequency of the circuit?

(c) At what time is the energy stored

(i) completely electrical (i.e. stored in the capacitor)? (ii) completely magnetic (i.e. stored in the inductor)?

(d) At what times is the total energy shared equally between the inductor and the capacitor?

(e) If a resistor is inserted in the circuit, how much energy is eventually dissipated as heat?

8.29 A coil of inductance 0.50 H and resistance 100 Ω is connected to a 240V, 50 Hz ac supply.

(a) What is the maximum current in the coil?

(b) What is the time lag between the voltage maximum and the current maximum?

8.30 Obtain the answers (a) to (b) above if the circuit is connected to a high frequency supply (240 V, 10 kHz). Hence explain the statement that at very high frequency, an inductor in a circuit nearly amounts to an open circuit. How does inductor behave in a dc circuit after the steady state?

8.31 A 100 μF capacitor in series with a 40 Ω resistance is connected to a 110V, 60 Hz supply.

(a) What is the maximum current in the circuit?

(b) What is the time lag between the current maximum and voltage maximum?

8.32 Obtain the answers to (a) and (b) above if the circuit is connected to a 110V, 12 kHz supply? Hence, explain the statement that a capacitor is a conductor at very high frequencies. Compare this behaviour with that of a capacitor in a dc circuit after the steady state.

Note: Pause here to think how a capacitor can conduct 'current' even there is nothing between its plates. You will get the answer in the next Chapter when you learn of 'displacement current'.

8.33 Keeping the source frequency equal to the resonating frequency of the series LCR circuit, if the three elements L, C and R are arranged in parallel, show that the total current in the parallel LCR circuit is minimum at this frequency. Obtain the current rms value in each branch of the circuit for the elements and source specified in 8.19 for this frequency.

8.34 A circuit containing a 80 mH inductor and a 60 μF capacitor in series is connected to a 230 V, 50 Hz supply. The resistance of the circuit is negligible.

(a) Obtain the current amplitude and rms values

(b) Obtain the rms values of potential drops across each element.

(c) What is the average power transferred to the inductor?

(d) What is the average power transferred to the capacitor?

(e) What is the total average power absorbed by the circuit? ['Average' implies 'averaged over one cycle']

8.35 Suppose the circuit in 8.34 has a resistance of 15 Ω. Obtain the average power transferred to each element of the circuit, and the total power absorbed.

8.36 A series LCR circuit with $L = 0.12$ H, $C = 480$ nF, $R = 23\Omega$ is connected to a 230 V variable frequency supply.

(a) What is the source frequency for which current amplitude is maximum. Obtain this maximum value.

(b) What is the source frequency for which average power absorbed by the circuit is maximum. Obtain the value of this maximum power.

(c) For which frequencies of the source is the power transferred to the circuit half the power at resonant frequency? What is the current amplitude at these frequencies?

(d) What is the Q-factor of the given circuit?

8.37 Obtain the resonant frequency and Q-factor of a series LCR circuit with $L = 3.0$ H, $C = 27\mu$F, and $R = 7.4\Omega$. It is desired to improve the sharpness of the resonance of the circuit by reducing its 'full width at half maximum' by a factor of 2. Suggest a suitable way.

8.38 Answer the following question:

(a) In any ac circuit, is the applied instantaneous voltage equal to the algebraic sum of the instantaneous voltages across the series elements of the circuit? Is the same true for rms voltage?

(b) Explain: voltage across L and C in series are 180^o out of phase, while for L and C in parallel, currents in L and C are 180^o out of phase.

(c) For circuits used for transporting electric power, a low power factor implies large power loss in transmission.

(d) Power factor can often be improved by the use of a capacitor of appropriate capacitance in the circuit.

(e) A capacitor is used in the primary circuit of an induction coil.

(f) An applied voltage signal consists of a superposition of a dc voltage and an ac voltage of high frequency. The circuit consists of an inductor and a capacitor in series. Show that the dc signal will appear across C and the ac signal across L.

(g) A choke coil in series with a lamp is connected to a dc line. The lamp is seen to shine brightly. Insertion of an iron care in the choke causes no change in the lamp's brightness. Predict the corresponding observations if the connection is to an ac line.

(h) A lamp is connected in series with a capacitor. Predict your observations for dc and ac connections. What happens in each if the capacity is reduced?

8.39 An ac generator consists of a coil of 50 turns and area 2.5 m^2 rotating at an angular speed of 60 rad s^{-1} in a uniform magnetic field $B = 0.30$ T between two fixed pole pieces. The resistance of the circuit including that of the coil is 500 Ω.

(a) What is the maximum current drawn from the generator?

(b) What is the flux through the coil when the current is zero? What is the flux when the current is maximum?

(c) Would the generator work if the coil were stationary and instead the pole pieces rotated together with the same speed as above?

8.40 A small dc motor operating at 200 V draws a current of 5.0 A at its full speed of 3000 rev/min. The resistance of the armature of the motor is 8.5 Ω. Determine the back emf of the motor. Obtain the power input, power output (mechanical) and the efficiency of the motor.

8.41 (a) Suppose the windings of the armature in the above dc motor cannot tolerate a current of more than 20A. What do you think will happen if the armature gets jammed and cannot rotate when the motor is connected to the supply?

(b) If the supply connections of the dc motor in 8.40 are removed and the motor is used as a generator by connecting the shaft of its armature to an external mechanical rotor of speed 3000 rev/min, how much emf will be generated?

8.42 A power transmission line feeds input power at 2300 V to a step down transformer with its primary windings having 4000 turns. What should be the number of turns in the secondary in order to get output power at 230 V?

8.43 At a hydroelectric power plant, the water pressure head is at a height of 300 m and the water flow available is 100 m^3s^{-1}. If the turbine generator efficiency is 60%, estimate the electric power available from the plant ($g = 9.8$ ms^{-2}).

8.44 A small town with a demand of 800 kW of electric power at 220 V is situated 15 km away from an electric plant generating power at 440 V. The resistance of the two wire line carrying power is 0.5 Ω per km. The town gets power from the line through a 4000-220 V step down transformer at a substation in the town.

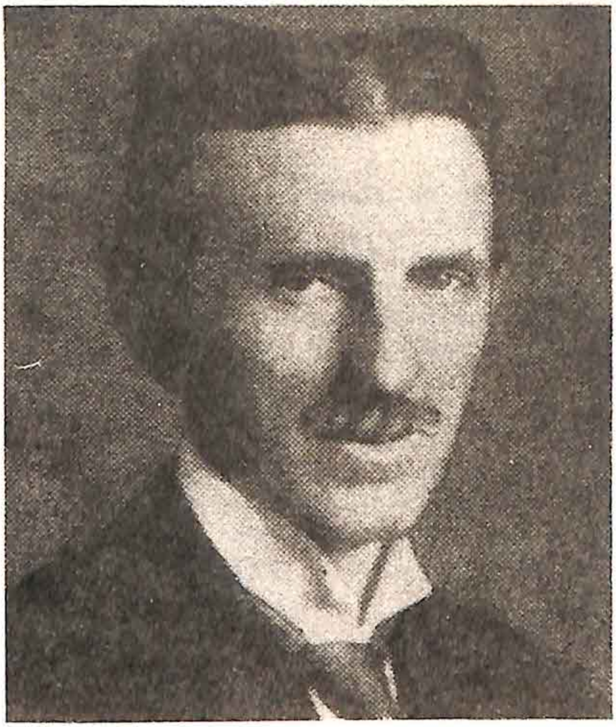
(a) Estimate the line power loss in

the form of heat.

(b) How much power must the plant supply, assuming there is negligible power loss due to leakage?

(c) Characterize the step up transformer at the plant.

8.45 Do the same exercise as above with the replacement of the earlier transformer by a 40,000-220 V step down transformer (Neglect, as before, leakage losses though this may not be a good assumption any longer because of the very high voltage transmission involved). Hence explain why high voltage transmission is preferred.



Tesla, Nicola (1836-1943) Yugosloy scientist, inventor and genius. He conceived the idea of the rotating magnetic field, which is the basis of practically all alternating current machinery, and which helped usher in the age of electric power. He also invented among other things the induction motor, the polyphase system of ac power, and the high frequency induction coil (the Tesla coil) used in radio and television sets and other electronic equipment. A master of mental visualization and spectacular demonstration, many of his observations and futuristic ideas remain unexploited. The SI unit of magnetic field is named in his honour.

CHAPTER 9

Electromagnetic waves

9.1 Introduction

We started this book with phenomena depending on electric charges at rest (Electrostatics). Then we began to look at what happens when charges move; in particular when there is a steady current (or flow) of charges. We found novel chemical, thermal and *magnetic* effects due to steady currents. The last two chapters have been concerned with the consequences of magnetic fields that are *not* steady, but change with time. Faraday and Henry found that such a field gives rise to an emf or an *electric field*. The final chapter in this story of the connections between electricity and magnetism was written by James Clerk Maxwell, the great British (Scottish) physicist, in the period 1855 to 1870.

Maxwell developed the ideas of Faraday on lines of force, and on electric and magnetic *fields*. He expressed the known laws of electricity and magnetism, namely Coulomb's (or Gauss's) law, Biot-Savart (or Ampere's) law and Faraday's law of electromagnetic induction in new precise, and unified ways in terms of electric and magnetic fields, and their sources. He noticed

an inconsistency in Ampere's law connecting electric currents and magnetic fields, and an asymmetry in the laws of electromagnetism - a term expressing the close connection between electricity and magnetism. He suggested the existence of an additional current, called by him the displacement current, which is such as to remove this inconsistency (Section 9.2). This also makes the laws of electricity and magnetism symmetrical. Maxwell then found that these new equations (now called Maxwell's equation) lead to the prediction that electric and magnetic fields dependent on time and space propagate as transverse waves, called electromagnetic waves. The velocity of these waves could be calculated from the theory which Maxwell had developed. It could also be estimated from electrical and magnetic measurements already made, using Maxwell's theory. The velocity turned out to be very close to that of light, known from optical measurements. "We can scarcely avoid the conclusion" said Maxwell in 1862 "that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena".

Maxwell's ideas on electromagnetic waves were directly and strikingly confirmed in 1887 by the experiments of Heinrich Hertz, who produced and detected electromagnetic waves (radio waves) in the laboratory. These were the beginnings of advances in science and technology that have changed the world, and the way we understand it.

In this Chapter, we present a descriptive account of electromagnetic waves. The displacement current is introduced in Section 9.2, and its consequences, namely electromagnetic waves, are outlined in the next section (Section 9.3). A brief history of the observation of electromagnetic waves follows (Section 9.4). The broad spectrum of electromagnetic waves, stretching from γ rays (wavelength $\sim 10^{-12}$m) to long radio waves (wavelength $\sim 10^{6}$m) is described in Section 9.5. Finally, the propagation of radio waves in the atmosphere is discussed in Section 9.6. A starred section at the end describes how messages are sent and received through radio waves.

9.2 Maxwell's displacement current: a new source of magnetic field

9.2.1 Inconsistency of Ampere's law

We have seen in Chapter 5 that an electric current produces a magnetic field around it. Maxwell showed that for logical consistency, a changing electric field *must also* produce a magnetic field. We shall present some arguments which make this reasonable. These arguments are similar to those used by Maxwell. We then discuss the meaning of this new source for a magnetic field, and mention some consequences.

In the last Chapter, we have seen that a capacitor *can be a circuit element* when an ac voltage, or more generally a time dependent voltage, is applied across it. On closer examination, this fact shows that Ampere's circuital law, Eq. (5.29) must be incomplete as stated. Consider a parallel plate capacitor C which is a part of a circuit through which a time dependent current $I(t)$ flows (Fig. 9.1).

Clearly, at a point such as P, in a region outside the parallel plate capacitor there is a nonzero magnetic field due to the current in the conducting wire connected to the capacitor. This can be seen, for example, using Ampere's circuital law (Fig. 9.2). Consider a plane circular loop of radius r whose plane is perpendicular to the direction of the current carrying wire, and which is centered symmetrically with respect to the wire (Fig. 9.2a). From symmetry, the magnetic field is

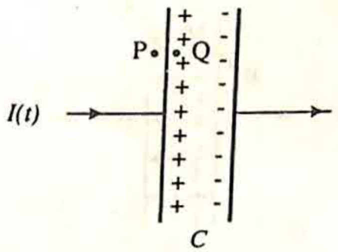


Figure 9.1: A parallel capacitor C, as a part of a circuit through which a time dependent electric current $I(t)$ flows. We are interested in the magnetic fields at point P and Q.

directed along the circumference of the circular loop and is the same in magnitude at all points on the loop so that if B is the magnitude of the field, the left hand side of Amperes circuital law equation is $B(2\pi r)$. The right hand side is μ_0 times the current through the plane of which the circle is the boundary. This current is I. So we have

$$B.2\pi r = \mu_0 I(t). \tag{9.1}$$

Now consider a different surface, which has the same boundary. This is a pot like surface, (Fig. 9.2b) which nowhere touches the current, but has its bottom between the capacitor plates; its mouth is the circular loop mentioned above. Another such surface is shaped like a tiffin box (without the lid) (Fig. 9.2c). On applying Ampere's law to such surfaces with the *same* perimeter, we find that the left hand side (magnetic field times the perimeter) has not changed, but the right hand side (current through the surface) is *zero* and *not I* since *no* current passes through the surfaces of Fig. 9.2b and 9.2c! So we have a flat contradiction; calculated one way, there is a magnetic field at a point such as P; calculated another way, there is none. Intuitively, we expect that there *should* be a magnetic field above the current carrying wire, at a point such as P. Since the contradiction arises from our use of Ampere's circuital law, this law must be missing something. The missing term must be such that one gets the same field no matter what surface is used.

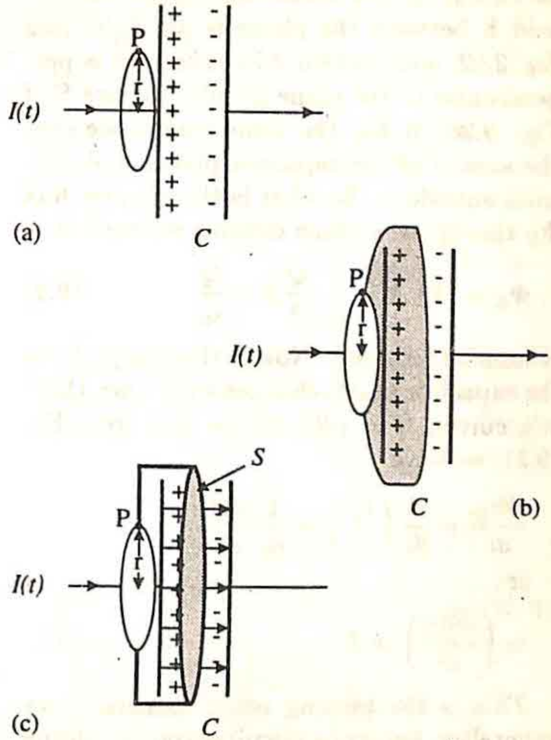


Figure 9.2: (a) The magnetic field at the point P can be calculated by imagining a circular loop passing through P (shown), which forms the boundary of the circular plane shown, and using Ampere's circuital law. (b) A pot shaped surface, with its circular rim passing through P. No conduction current passes through this pot shaped surface. (c) A tiffin box shaped surface with its circular rim the same as that in Fig. (9.2b), but with a flat circular bottom S between the capacitor plates. The uniform electric field $\mathbf{E}$ between the capacitor plates is indicated.

9.2.2 Displacement current

We can actually guess the missing term by looking carefully at Fig. 9.2c. Is there any-

thing passing through the plane circular surface S *between* the plates of the capacitor? Yes, of course, the electric field! If the plates of the capacitor have an area A, and a total charge Q, the magnitude of the electric, field $\mathbf{E}$ between the plates is $(Q/A)/\epsilon_0$ (see Eq. 2.62, and Section 2.7). The field is perpendicular to the plane circular surface S of Fig. 9.2c. It has the same magnitude over the area A of the capacitor plates, and vanishes outside it. So what is the *electric* flux Φ_E through the plane circular surface? It is

$$\Phi_E = |\mathbf{E}|.A = \frac{1}{\epsilon_0}\frac{Q}{A}A = \frac{Q}{\epsilon_0} \tag{9.2}$$

(Gauss's theorem!). Now if the charge Q on the capacitor plates changes with time, there is a current $I = (dQ/dt)$, so that from Eq. (9.2), we have

$$\frac{d\Phi_E}{dt} = \frac{d}{dt}\left(\frac{Q}{\epsilon_0}\right) = \frac{1}{\epsilon_0}\frac{dQ}{dt}$$

or

$$\epsilon_0\left(\frac{d\Phi_E}{dt}\right) = I. \tag{9.3}$$

This is the missing term; namely, if we generalize Ampere's circuital law by adding to the total current carried by conductors through the surface, another term which is ϵ_0 times the rate of change of electric flux through the same surface, the *total* has the same value I for all surfaces. If this is done, there is no contradiction in say, the value of B obtained anywhere using the generalized Ampere's law. B at the point P is nonzero no matter which surface is used for calculating it. B at a point P outside the plates (Fig. 9.1a) is nearly the same as at a point Q just inside, as it should be. The first current, carried by conductors and due to flow of charges, is called *conduction current.* The second, namely (Eq. (9.3)), is a new term, and is due to changing electric field or electric *displacement* (an old term still used sometimes). It is, therefore, called *displacement current* or Maxwell's displacement current. Fig. 9.3 shows the electric and magnetic fields inside the parallel plate capacitor discussed above.

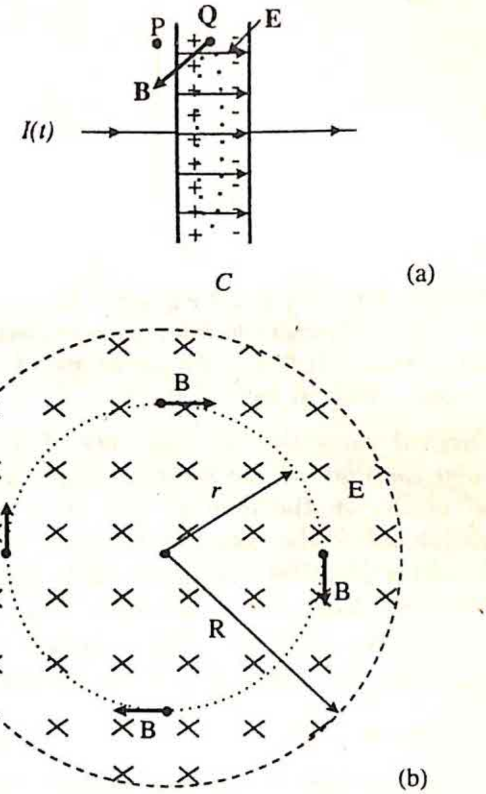


Figure 9.3: (a) The electric and magnetic fields **E** and **B** between the capacitor plates, at the point Q. (b) A cross sectional view of Fig 9.3a

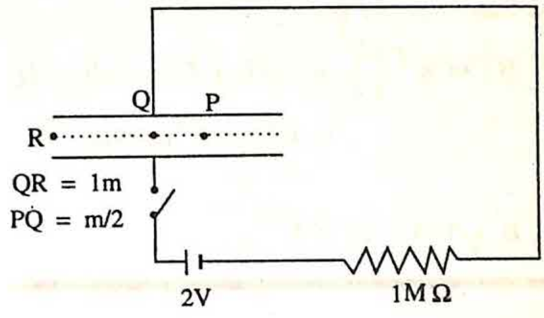
The generalization made by Maxwell then is the following. The source of a magnetic field is not *just* the conduction electric current due to flowing charges, but also the time rate of change of electric field. More precisely, the total current I is the sum of the conduction current denoted by I_c, and the displacement current denoted by $I_d (= \epsilon_0 (d\Phi_E/dt))$. We have,

$$I = I_c + I_d = I_c + \epsilon_0\frac{d\Phi_E}{dt} \tag{9.4}$$

The generalized (and correct) Ampere's circuital law has the same form as Eq. (5.29), with one difference: "the *total current* passing through any surface of which the closed circuit is the perimeter" is the sum of the conduction current and the displacement current (Eq. (9.4)). The displacement current is ϵ_0 times the rate of change of flux of electric field through the same surface as that through which the flow of conduction current is calculated.

In all respects, the displacement current has the same physical effects as the conduction current. In some cases, e.g. steady electric fields in a conducting wire, the displacement current may be zero since the electric field **E** does not change with time. In other cases, for example, the charging capacitor above, both conduction and displacement currents may be present in different regions of space. In some other cases, they may both be present in the same region of space. Most interestingly, there may be large regions of space where there is *no* conduction current, but there is only a displacement current due to time varying electric fields. In such a region, we expect a magnetic field though there is no (conduction) current source nearby! The prediction of such a displacement current is easily verified experimentally, for example by measuring the magnetic field between the plates of the capacitor in Fig. 9.3. The magnetic field (say at Q) can be measured and is seen to be the same as that just outside (at P).

Example 9.1 : A parallel plate capacitor with circular plates of radius 1 m has a capacitance of 1 nF. At $t = 0$, it is connected for charging in series with a resistor $R = 1\,\mathrm{M}\Omega$ across a 2V battery. Calculate the magnetic field at a point P, in between the plates and halfway between the centre and the periphery of the plates, after 10^{-3}s.



Answer : The time constant of the charging CR circuit is $\tau = CR = 10^{-3}$s. The charge on the plate at time t, then, is according to the formula given in the last chapter.

$$\begin{aligned} q(t) &= CV[1 - \exp(-t/\tau)] \\ &= 2 \times 10^{-9}[1 - \exp(-t/10^{-3})]. \end{aligned}$$

The electric field in between the plates at time t is:

$$\begin{aligned} E &= \frac{q(t)}{\epsilon_0 A} = \frac{q(t)}{\pi\epsilon_0}; A = \pi(1)^2 \\ &= \text{area of the plates.} \end{aligned}$$

Consider now a circular loop of radius (1/2) m parallel to the plates passing through P. The magnetic field B at all points on this loop is along the loop and of the same value.

The flux Φ_E through this loop is

$$\begin{aligned} \Phi_E &= E \times \text{area of the loop} \\ &= E \times \pi \times \left(\frac{1}{2}\right)^2 = \frac{\pi E}{4} = \frac{q}{4\epsilon_0} \end{aligned}$$

The displacement current.

$$\begin{aligned} I_d &= \epsilon_0 \frac{d\Phi_E}{dt} \\ &= \frac{1}{4}\frac{dq}{dt} = 0.5 \times 10^{-6} \exp(-1) \end{aligned}$$

at $t = 10^{-3}$s. Now, applying Ampere's law to this loop, we get

$$B \cdot 2\pi \times \left(\frac{1}{2}\right) = \mu_0(I_c + I_d) = \mu_0(0 + I_d)$$
$$= 0.5 \times 10^{-6}\mu_0 \exp(-1)$$

or

$$B = 0.74 \times 10^{-13}\,\mathrm{T}$$

9.2.3 Consequences of displacement current

The displacement current has (literally) far reaching consequences. One thing we immediately notice is that the laws of electricity and magnetism are now more symmetrical.[1] Faraday's law of induction states that there is an induced emf *equal to the rate of* change of magnetic flux. Now since the emf between two points 1 and 2 is the work done per unit charge in taking it from 1 to 2, the existence of an emf implies the existence of an electric field. So, we can loosely rephrase Faraday's law of electromagnetic induction by saying that a *magnetic field* changing with time gives rise to an *electric field.*

That an *electric field* changing with time gives rise to a *magnetic field* is the symmetrical counterpart, and is a consequence of the displacement current being a source of a magnetic field. Thus time dependent electric and magnetic fields give rise to each other! This statement can be made quantitative. One quantitative relation is Eq. (7.3) i.e. Faraday's law of electromagnetic induction. Another is Ampere's circuital law, with the current being the total current as in Eq. (9.4). One very important consequence of this symmetry is electromagnetic waves, as we discuss qualitatively in the next Section.

[1]They are still not perfectly symmetrical; there are no known sources of magnetic field (magnetic monopoles) analogous to electric charges which are sources of electric field.

9.3 Electromagnetic waves

9.3.1 Existence of electromagnetic waves

Maxwell showed that the complete set of laws of electricity and magnetism (now called Maxwell's equations) imply the existence of electric and magnetic fields **E** and **B** that propagate through space with the velocity of light. He also showed that the waves are transverse. We shall not prove these facts here, but will try to make them plausible, and describe the nature of electromagnetic waves.

If, in a region of space, there is a flux of magnetic field varying with time, it gives rise to an emf (Faraday's law of electromagnetic induction). Applying this law to a small region, say a small square perpendicular to the direction of the magnetic field, it can be shown that the electric fields along the two parallel sides of the square are not the same. Thus a time dependent magnetic field gives rise to an electric field that varies with position. Now suppose that the electric field depends on *time* as well. Then from Maxwell's generalization of Ampere's circuital law, such an electric field produces a magnetic field (which can again be shown to vary in space). So, we find that electric and magnetic fields that depend on space and time produce and sustain each other. A simple form of such a continuing change is a wave. In a plane wave, for example, the electric and magnetic fields vary sinusoidally with distance at a given time, and with time at a given point (see below for an explicit form). We will guess the speed of propagation of these waves from dimensional arguments, and also argue that electric and magnetic fields in an electromag-

netic wave are perpendicular to each other. We then describe the electromagnetic wave precisely, in terms of the actual electric and magnetic fields which are perpendicular to each other and to the direction of propagation. We also mention quantities such as the energy and pressure of electromagnetic waves. In the next section we describe how electromagnetic waves can be produced.

9.3.2 Velocity of electromagnetic waves

The velocity of propagation of electromagnetic waves can be guessed purely dimensionally, without solving Maxwell's equations for electric and magnetic fields. If E is the electric field, B the magnetic field, and L and T are dimensional symbols for length and time, Faraday's law of electromagnetic induction can be written dimensionally as

$$[EL] = [BL^2T^{-1}] \tag{9.5a}$$

or

$$[EL^{-1}] = [BT^{-1}] \tag{9.5b}$$

Similarly, the Ampere's circuital law, in the absence of a conduction current or presence *only* of a displacement current schematically becomes

$$[BL] = [(\mu_0\epsilon_0)EL^2T^{-1}] \tag{9.6a}$$

or

$$[BL^{-1}] = [(\mu_0\epsilon_0)ET^{-1}] \tag{9.6b}$$

Equations (9.5) and (9.6) describe, very crudely, how electric and magnetic fields depending on space and time produce each other. We can use Eq. (9.5b) to eliminate B from Eq. (9.6b) which then becomes

$$[EL^{-2}] = [(\mu_0\epsilon_0)ET^{-2}]$$

or

$$[E] = [(\mu_0\epsilon_0)(L/T)^2E]. \tag{9.7}$$

This clearly suggests that if we have a time and space dependent electric field, it propagates with a velocity $(L/T) \simeq (\mu_0\epsilon_0)^{-1/2}$. In a dimensional estimate we cannot get the numerical factor correct, but it turns out that indeed

$$v_{em} = c = (1/\sqrt{\mu_0\epsilon_0}) \tag{9.8}$$

The magnetic field also propagates with the same velocity, namely, velocity of light in free space.

9.3.3 Nature of electromagnetic waves

A real calculation shows that electric and magnetic fields in an electromagnetic wave are perpendicular to each other, *and* to the direction of propagation. The former appears reasonable, say from our discussion of the displacement current. Consider Fig. 9.3. The electric field inside the plates of the capacitor is directed perpendicular to the plates. The magnetic field this gives rise to via the displacement current is along the perimeter of a circle parallel to the capacitor plane. So **B** and **E** are perpendicular in this case. This is a general feature.

In Fig.(9.4), we show a typical example of a plane electromagnetic wave propagating along the z direction (the fields are shown as a function of the z coordinate, at a given time t). The electric field E_x is along the x axis, and varies sinusoidally with z, at a given time. The magnetic field B_y is along the y axis, and again varies sinusoidally with z. The electric and magnetic fields E_x and B_y are perpendicular to each other, and to the direction z of propagation. We can write E_x and B_y as follows:

$$E_x = E_{x_0}\sin(kz - \omega t) \tag{9.9a}$$

$$B_y = B_{y_0}\sin(kz - \omega t) \tag{9.9b}$$

Here k is related to the wave length λ of the wave by the usual equation

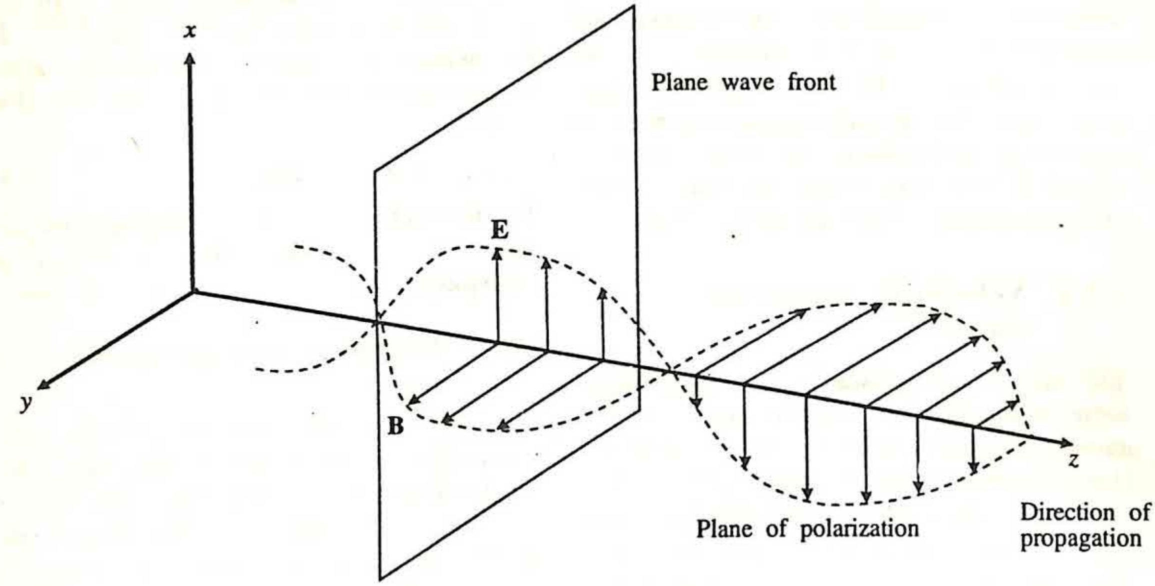


Figure 9.4: A linearly polarized electromagnetic wave, propagating in the z direction with the oscillating electric field **E** along the x direction and the oscillating magnetic field **B** along the y direction.

$$k = \frac{2\pi}{\lambda}, \tag{9.10}$$

and ω is the angular frequency. k is actually the magnitude of the wave vector (or propagation vector) **k**. Its direction describes the direction of propagation of the wave, and its magnitude k is related to the wavelength as $(2\pi/\lambda)$ (Eq. (9.10)). The speed of propagation of this wave is (ω/k). When the forms Eq. (9.9) for E_x and B_y are substituted into the laws of electricity and magnetism, namely Maxwell's equations, one finds that

$$\omega = ck, \text{ where } c = \frac{1}{\sqrt{\mu_0\epsilon_0}} \tag{9.11}$$

This is what was guessed earlier, in Eq. (9.8).

The relation $\omega = ck$ is the standard one for waves (see eg. Chapter 13 of class XI Physics textbook). This relation is often written in terms of frequency $\nu(=\omega/2\pi)$ and wavelength $\lambda(=2\pi/k)$ as

$$2\pi\nu = c\left(\frac{2\pi}{\lambda}\right)$$

or

$$\nu\lambda = c \tag{9.12}$$

It is also seen from Maxwell's equations that electric and magnetic fields in an electromagnetic wave are related; namely

$$B_{y_0} = (E_{x_0}/c). \tag{9.13}$$

The dimensional relation Eq. (9.5) is consistent with this result.

We remark on some features of electromagnetic waves. As described above, they are self sustaining oscillations of electric and magnetic fields in free space, or vacuum. They differ from all the other waves we

have studied so far in this respect, that *no material medium* is involved in the vibrations of the electric and magnetic fields. Sound waves in air are longitudinal waves of air compression and rarefaction. Transverse waves on the surface of water consist of water moving up and down as the wave spreads horizontally and radially onwards. Transverse elastic (sound) waves can also propagate in a solid, which is rigid and that resists shearing. Scientists in the nineteenth century were much used to this mechanical picture and thought that there must be a subtle medium, pervading all space and all matter, which responds to electric and magnetic fields much as any elastic medium does. They called this medium 'ether'. They were so convinced of the reality of this medium, that there is even a novel called the 'Poison Belt' by Arthur Conan Doyle (the creator of the famous detective Sherlock Holmes) where the solar system is supposed to pass temporarily through a poisonous region of ether! We now accept that no such physical medium is needed. Electric and magnetic fields, oscillating in space and time, can sustain each other in vacuum. The ether idea leads to predictions which are not confirmed experimentally.

★ The experiments referred to above are the famous experiments of Michelson and Morley which attempted to find the change in the velocity of light as the earth moves through 'ether' with different velocities during the year. No such change was found, and this led Einstein to abandon the notion of ether altogether as experimentally untenable and theoretically unncessary, and develop a new theory of space and time. This theory, called the special theory of relativity, is amply confirmed by experiments. The velocity of light plays a special role in this theory eg. as a limiting velocity. Einstein also developed a new theory of gravitation consistent with the special theory of relativity. This is called the general theory of relativity, and is again found experimentally to be correct. Newton's theory of gravitation (you have studied in class XI) is what Einstein's theory reduces to in the limit that the gravitating systems move with speeds much slower than the speed of light. ★

But what if a material medium is actually there? We know that light, an electromagnetic wave, does propagate through glass for example. We have seen earlier that the total electric and magnetic fields inside a medium are described in terms of a permittivity ϵ and a magnetic permeability μ. (These describe the factors by which the total fields differ from the external fields; see Sections 2.8 and 6.4 respectively for details). These replace ϵ_0 and μ_0 in the description of electric and magnetic fields in Maxwell's equations so that using the same ideas as before, one can argue that in a material medium of permittivity ϵ and magnetic permeability μ, the velocity of light

$$c = \frac{1}{\sqrt{\mu\epsilon}} \tag{9.14}$$

The velocity of light depends on electric and magnetic properties of the medium. We shall see in the next Chapter that the *refractive index* of one medium with respect to the other is equal to the ratio of velocities of light in the two media.

The velocity of electromagnetic waves in free space or vacuum is an important fundamental constant. It has been shown by experiments on em waves of different wavelength that this velocity is the same (independent of wavelength) to within a few metres per second, out of a value of order 3×10^8 m/s. You will read about measurements of the speed of light in Chapter 11. The speed of microwaves and radiowaves is measured by very different meth-

ods, in which the frequency ν and the wavelength λ are determined separately. The belief in the constancy of the velocity of em waves is so strongly supported by experiment and the actual value is so well known now that this is taken as the standard of *length*. Namely, the metre is now *defined* as the distance travelled by light in vacuum in a time $(1/c)$ seconds $= (2.99792458 \times 10^8)^{-1}$ seconds. This has come about for the following reason. The basic unit of time can be defined very accurately in terms of some atomic frequency, i.e. frequency of light emitted by an atom in a particular process. The basic unit of length is harder to define as accurately in a direct way. Earlier measurements of c, using earlier units of length (metre rods $\cdots$) converged to a value of about 2.9979246×10^8m/s. Since c is such a strongly fixed number, unit of length can be defined in terms of c and the unit of time!

Another interesting feature of electromagnetic waves is that no physical source of the oscillating electric and magnetic fields is in evidence. The situation is similar to that of a spring, which is a simple harmonic oscillator. If it is pulled to one side and let go, it will continue to oscillate forever in the absence of friction. So, one question is, what is the equivalent of pulling the spring? How can we produce electromagnetic waves, how can we sustain them, and how do we detect them? This is a large subject, now mostly a branch of electrical engineering, but we will mention a few essential things below (section 9.4). *Essentially, oscillating charges produce electromagnetic waves.* But before that, we briefly go into the questions of energy and pressure associated with an electromagnetic wave.

We have seen in chapter 2 (Eq. (2.85) and later discussion) that in a region of free space with electric field $\mathbf{E}$, there is an energy density $(\epsilon_0 E^2/2)$. Similarly, associated with a magnetic field $\mathbf{B}$ is a magnetic energy density $(B^2/2\mu_0)$. Since in an electromagnetic wave, one has both electric and magnetic fields, there is a nonzero energy density associated with it. Now consider a plane perpendicular to the direction of propagation of the electromagnetic wave. If there are, on this plane, electric charges, they will be set and sustained in motion by the electric and magnetic fields of the electromagnetic wave. The charges thus acquire energy and momentum from the wave. This just illustrates the fact that an electromagnetic wave (like other waves) carries energy and momentum. We do not calculate these here. Since it carries momentum, an electromagnetic wave also exerts pressure, called radiation pressure.

Example 9.2: A plane light wave in the visible region is moving along the z-direction. The frequency of the wave is 0.5×10^{15} Hz and the electric field at any point is varying sinusoidally with time with an amplitude of $1\ \mathrm{Vm}^{-1}$. Calculate the average values of the energy densities of the electric and magnetic fields.

Answer: The electric field is at right angles to the z-direction, say in the x-direction. We can then write:

$$\begin{aligned} E_y &= E_z = 0. \\ E_x &= (1\mathrm{Vm}^{-1}) \sin\left[2\pi\left(\frac{z}{\lambda} - \nu t\right)\right] \end{aligned}$$

where $\nu = 0.5 \times 10^{15}$ Hz and λ is the wavelength. The energy density of the electric field is then

$$\begin{aligned} u_E &= \frac{1}{2}\epsilon_0 |\mathbf{E}|^2 \\ &= \frac{\epsilon_0}{2}(1)^2 \sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right)\right]. \end{aligned}$$

The function $\sin^2[2\pi((z/\lambda) - \nu t)]$ at any z oscillates with time between 0 and 1, with

an average value 1/2. Hence, the average u_E is

$$\begin{aligned}\overline{u}_E &= \frac{\epsilon_0}{2}(1)^2\frac{1}{2} = \frac{\epsilon_0}{4} \\ &\simeq 2.21 \times 10^{-12}\mathrm{Jm}^{-3}.\end{aligned}$$

Corresponding to an electric field in the x-direction, we will have a magnetic field in the y-direction as given in the text:

$$\begin{aligned}B_x &= B_z = 0 \\ B_y &= B_y^{(0)} \sin\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right] \\ B_y^{(0)} &= \mu_0\epsilon_0\frac{\omega}{k}E_x^{(0)}\end{aligned}$$

where $\omega = 2\pi\nu, k = \frac{2\pi}{\lambda}$ and $E_x^{(0)}$ is the amplitude of E_x given as 1 Vm^{-1}. Hence

$$\begin{aligned}B_y^2 &= (B_y^{(0)})^2 \sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right] \\ &= \left(\mu_0\epsilon_0\frac{\omega}{k}\right)^2 (E_x^{(0)})^2 \times \\ &\quad \sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right].\end{aligned}$$

The magnetic energy density is

$$\begin{aligned}\overline{u}_B &= \frac{1}{2}\frac{|\mathbf{B}|^2}{\mu_0} \\ &= \frac{1}{2}\frac{\mu_0^2\epsilon_0^2(\omega/k)^2(E_x^{(0)})^2}{\mu_0} \times \\ &\quad \sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right] \\ &= \frac{1}{2}\epsilon_0(E_x^{(0)})^2 \sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right].\end{aligned}$$

Since $\omega/k = c = 1\sqrt{\mu_0\epsilon_0}$. The average value of $\sin^2\left[2\pi\left(\frac{z}{\lambda} - \nu t\right) + \varphi\right]$ at any point is 1/2 and hence the average value of u_B is

$$\frac{1}{2}\epsilon_0(1)^2\frac{1}{2} = \frac{\epsilon_0}{4},$$

the same as the average value of u_E.

★ The intensity of any wave is the power flowing across an area (perpendicular to the direction of propagation of the wave), per unit area. This can be shown to be the product of the energy density of the wave and its speed. This equality is reasonable, since if the speed is large, in a given time a lot of energy will be deposited in the unit area mentioned above. Thus for the electromagnetic wave, the intensity

$$I = \left(\frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0}\right)c \tag{9.15}$$

Since E and B both oscillate sinusoidally with time, the average intensity averaged over a cycle is

$$I = \left(\epsilon_0\frac{E_0^2}{4} + \frac{B_0^2}{4\mu_0}\right)c \tag{9.16}$$

where E_0 and B_0 are the maximum values of the fields, the E_{x_0} and B_{y_0} of Eq. (9.9). Also, since E_{x_0} and B_{x_0} are related e.g. Eq. (9.13), the intensity I can be written as

$$I = \epsilon_0\frac{E_0^2}{2}c = \frac{1}{2}\frac{E_0B_0}{\mu} \tag{9.17}$$

Thus the intensity of an electromagnetic wave is proportional to the square of the electric (or magnetic) field, or conversely, the size of the electric (or magnetic) field in an electromagnetic wave is the square root of its intensity.

The electromagnetic wave also carries momentum. It can be shown that the momentum is actually (1/c) times the energy. Since energy per unit time per unit area is intensity, the momentum per unit time per unit area, or force per unit area (pressure!), is related to intensity as momentum is related to energy. We have pressure P_r due to electromagnetic radiation to be given by

$$P_r = \frac{I}{c} \tag{9.18}$$

This pressure, called *radiation pressure*, is real, of course. Among other things, this is the reason why tails of comets point away from the sun! ★

9.4 Production and observation of electromagnetic waves

9.4.1 Accelerated charge is the source of electromagnetic waves

At the time Maxwell realized that electromagnetic *waves* follow from the laws of electricity and magnetism, the only waves known to propagate in space were light waves. He also noted (as mentioned in the introduction) that the velocity of electromagnetic waves from known electrical and magnetic measurements (due to G. Kohlrausch and W. Weber in 1856) is the same as that of light. This, and other experimental results such as the effect of magnetic fields on light waves (discovered by Faraday in 1844) clearly and strongly suggest that light is an electromagnetic wave, with wavelengths in the region 4000Å to 8000Å. However, the equations of Maxwell put no restriction on the wavelength of electromagnetic waves; they should exist at all wavelengths. How can one produce and detect electromagnetic waves of other wavelengths? We first give a qualitative general answer, and then describe the production and observation of radio waves (electromagnetic waves with wavelength λ in the range 1 cm to 1 km).

We know that a stationary charge produces an electric field, and that a uniformly moving charge or a steady current produces a steady or stationary magnetic field. An *accelerated* charge then produces a magnetic field which changes with time and of course depends on space. As has been discussed in earlier sections, an electromagnetic wave is associated with such a magnetic field, i.e. one which is dependent on space and time. Thus, an *accelerated electric charge* is the source of electromagnetic radiation. A common realization of such an acceleration is a simple harmonic oscillation. A charge oscillating harmonically with a frequency ν is the source of electromagnetic waves of the same frequency; it produces electric and magnetic fields at any point in space which oscillate with frequency ν. This is somewhat like a stick dipped in water, and moving rhythmically up and down at a particular frequency. The frequency of the surface water wave produced is that of the up and down motion of the stick (forced oscillations).

Oscillating electric charges come in all sizes and shapes. Electrons in atoms, bound to the nucleus, oscillate. This gives rise to visible light.

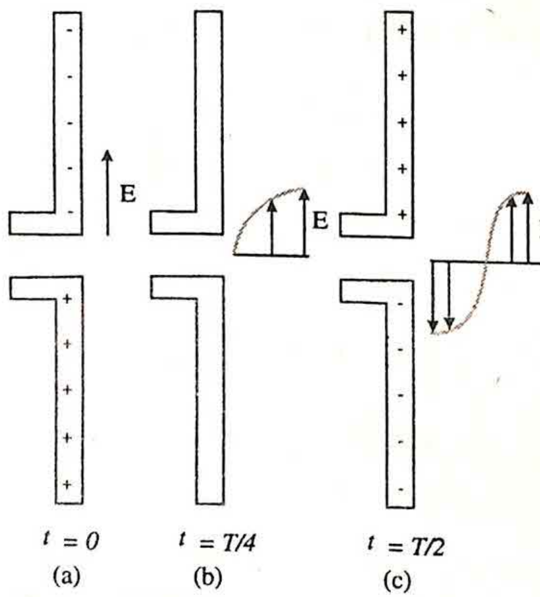


Figure 9.5: An electric dipole antenna. The electric charges on the antenna, and the electric field produced are shown for $t = 0, t = T/4$ and $t = T/2$ where T is the time period of the electromagnetic wave. The magnetic field is not shown.

9.4.2 Production and detection of radio waves

Atomic oscillators or oscillating electric dipoles have a size of order 1Å or 10^{-10}m.

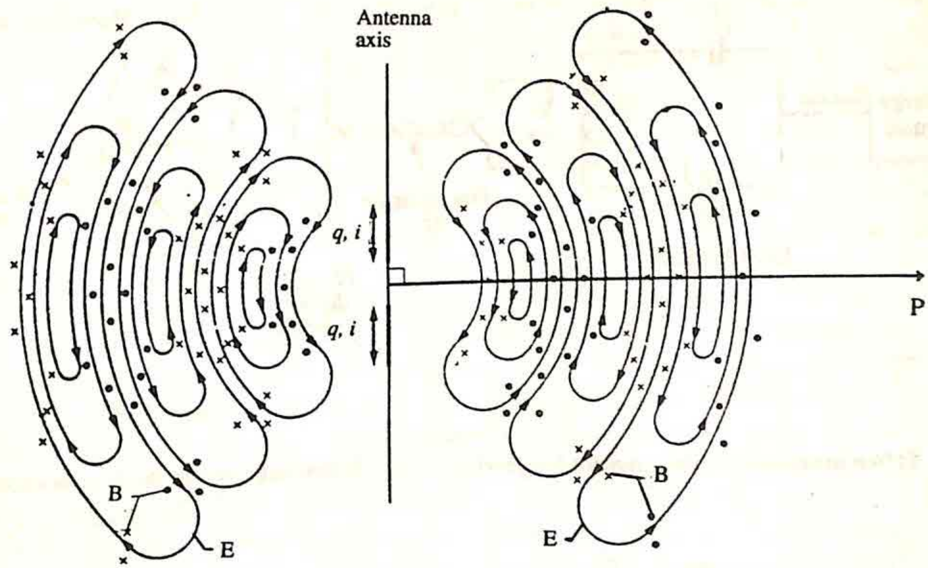


Figure 9.6: The electric and magnetic field lines due to a dipole antenna at a given instant of time. The **E** fields is shown by lines, and the **B** field by dots and crosses.

The visible electromagnetic waves they produce have a wavelength ranging from 7000Å (7×10^{-7}m; red) to 4000Å (4×10^{-7} m; violet) at frequencies in the range 4×10^{14} Hz (red) to 7×10^{14} Hz (violet). It is clear from this that oscillating electric dipoles of much larger size are needed to produce electromagnetic waves of longer wavelength. For example, even if the same ratio between the size of the source and the size of the wavelength is maintained, i.e. $\sim 2 \times 10^{-4}$, one needs an oscillating electric dipole of about 2 cm size to produce electromagnetic waves of wavelength $\sim$ 100 m (frequency 3×10^6 Hz). Such a dipole can consist of two bent conducting rods fed by an ac generator! This produces an alternating current as well as charge along the conducting rods, called a dipole antenna. This dipole antenna produces electromagnetic waves at the frequency of the ac generator (Figs. 9.5 and 9.6). Such an arrangement was first thought of by Hertz, and is also called a Hertzian dipole. We show in Fig. (9.6) the electric and magnetic fields produced by such an oscillating dipole at different times, and at different points in space. You can see the electromagnetic wave spreading out (radiating out) from the dipole antenna.

From conservation of energy, it is clear that the ac generator has to deliver power to the antenna at the rate at which it radiates electromagnetic wave energy. Though we have described above oscillating dipoles that are small in size compared to the wavelength of the em (short form for electromagnetic) radiation they emit, it turns out that the most efficient antennas are those which have a size comparable to the wavelength of the electromagnetic waves they produce. A schematic diagram of an arrangement for generating electromagnetic waves is given in Fig. 9.7.

We briefly touch upon the basic principles

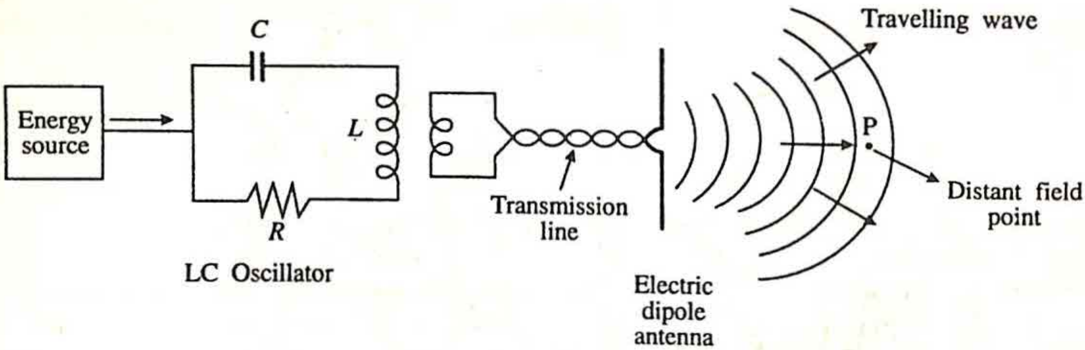


Figure 9.7: An arrangement (schematic) for producing electromagnetic waves in the shortwave radio region.

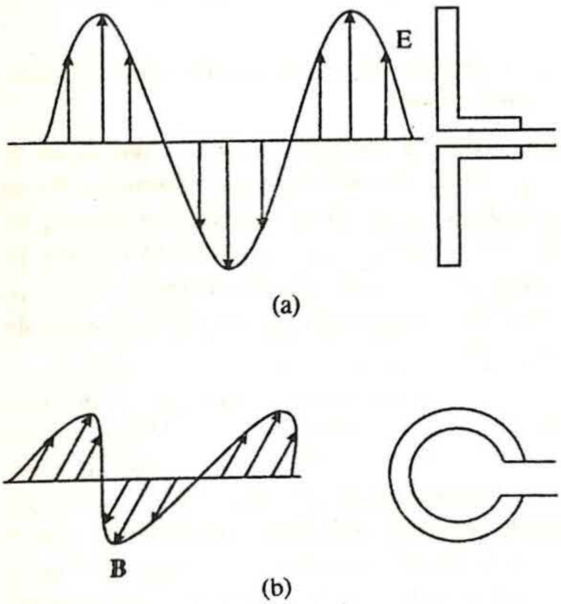


Figure 9.8: (a) An electric dipole antenna for detecting electromagnetic radiation. The alternating current electric field of the radiation produces an alternating current in the antenna. (b) A loop antenna detector of electromagnetic radiation. The changing magnetic flux in the loop due to the incident radiation induces an alternating emf in the loop.

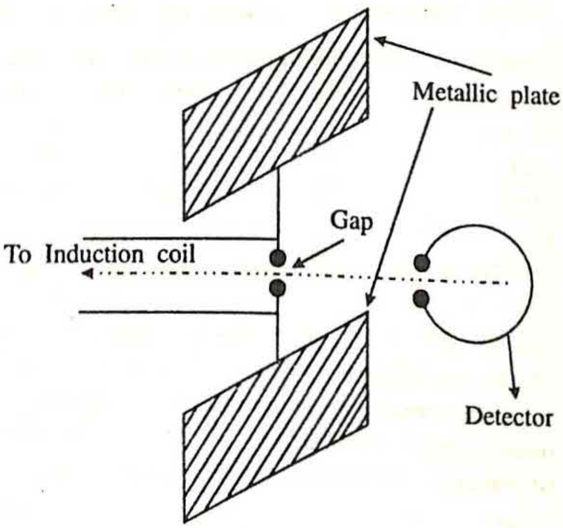


Figure 9.9: Schematic diagram of Hertz's experimental set up. The metal plates are charged to a high voltage by an induction coil. When the voltage is sufficiently high the plates get discharged by sparking across the narrow gap.

involved in the detection of radio or television waves (electromagnetic waves with $\lambda \sim$ 20 cm to 200 m, say). Suppose a dipole antenna is oriented such that the electric field of the incident electromagnetic wave is parallel to it (Fig. 9.8a). The oscillating electric field causes the free electrons in the antenna to oscillate with the same frequency as itself (forced oscillations) and thus sets up an alternating current. This alternating current can be detected in many ways. One such is to connect the antenna to an LC circuit. If the frequency of the LC circuit is varied (turned), the largest current will flow when its frequency $\nu = (2\pi\sqrt{LC})^{-1}$ is the same as that of the electromagnetic wave (resonance). Electromagnetic waves can also be detected by a loop antenna (Fig. 9.8b). This is oriented perpendicular to the magnetic field of the incident em wave, so that the changing magnetic flux induces an alternating current in the loop which can be detected via a resonant circuit. The size of the antenna is comparable to that of the wavelength of the em wave to be detected.

You must have seen advanced versions of such antennas or ariels on many housetops, where they receive radio and TV signals broadcast from all over the world! Electromagnetic waves of frequencies in the visible light range can be detected by the eye or by photographic film. Both of these are mainly sensitive to the electric field of the electromagnetic wave.

9.4.3 History of observation of electromagnetic waves

We now briefly summarize some aspects of the discovery of electromagnetic waves in the metre to millimeter range. As mentioned above, Hertz realized in 1887 that electromagnetic waves in this wavelength range could be produced in the laboratory using oscillating dipole antennas and could be detected similarly. Fig. (9.9) is a schematic representation of Hertz's set up. The two metal sheets ending in a dipole antenna are connected to a source of high voltage (Hertz used the induction coil, which produces high voltage by electromagnetic induction). The voltage is high enough so that the air in the small gap between the plates gets ionized and provides a path for discharge of the plates. The plate arrangement has very low inductance L and capacitance C so that the resonant angular frequency $\omega = (1/\sqrt{LC})$ is very high, and very high frequency oscillations of charges on the plates will result.

With the kind of arrangement just outlined, Hertz was able to produce electromagnetic waves of wavelength around 6m. The detector is also shown in our schematic diagram Fig. 9.9. It is held in a position such that the magnetic field produced by the oscillating current is perpendicular to the plane of the coil. The resultant electric field, induced by the oscillating magnetic field causes sparks to appear at the narrow gap. Hertz showed by direct experiments that the electromagnetic waves he produced have all the properties of light, e.g. same speed of propagation, reflection, refraction, diffraction, interference and polarization. This conclusively demonstrated the existence of the same kind of electromagnetic waves over a wide range of wavelengths, from a few metres to about 4×10^{-7} metres. The SI unit of frequency of oscillation is named Hertz (abbreviated Hz) in honour of this pioneer.

The successful demonstration of electromagnetic waves created a sensation and sparked off other important achievements. Two important achievements in this connection deserve mention. Eleven years after Hertz, Jagadish Chandra Bose, working at Calcutta, succeeded in producing and

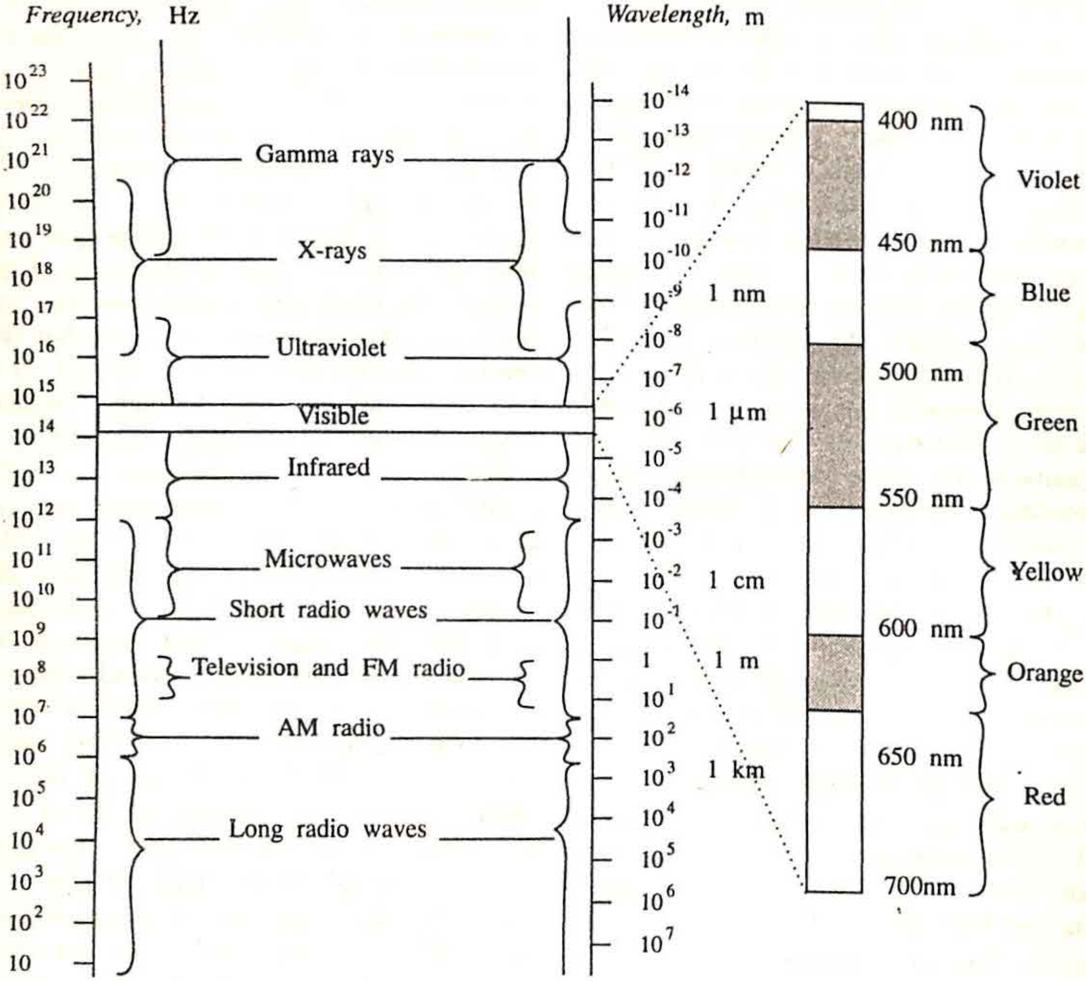


Figure 9.10: The electromagnetic spectrum, with common names for various parts of it. The various regions do not have sharply defined boundaries.

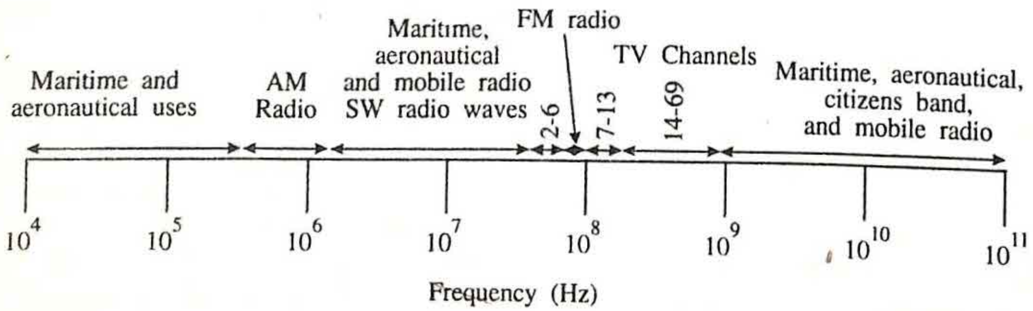


Figure 9.11: Frequency bands in radio wave communication.

Table 9.1: Radio frequency bands.

Band	Frequency range	Wavelength range
Extremely low frequency (ELF)	< 3 kHz	> 100 km
Very low frequency (VLF)	3-30 Hz	10-100 km
Low frequency (LF)	30-300 kHz	1-10 km
Medium frequency (MF)	300 kHz-3 MHz	100 m-1 km
High frequency (HF)	3-30 MHz	10-100 m
Very high frequency (VHF)	30-300 MHz	1-10 m
Ultra high frequency (UHF)	300 MHz-3 GHz	10 cm-1 m
Superhigh frequency (SHF)	3-30 GHz	1-10 cm
Extremely high frequency (EHF)	30-300 GHz	1 mm-1 cm

observing electromagnetic waves of much shorter wavelength (25 mm to 5 mm). His experiments, like Hertz's, were confined to the laboratory. At around the same time Guglielmo Marconi, in Italy, followed Hertz's work but succeeded in transmitting electromagnetic waves over distances of many miles. The oscillator used by Marconi was similar but different in detail compared to Hertz's; one metal plate was atop a pole while the other was on the surface of the earth. Marconi's experiments mark the beginning of the field of communications using electromagnetic waves.

9.5 Light and the spectrum of electromagnetic radiation

Beginning with the demonstration by Hertz of the existence of electromagnetic waves, belief in Maxwell's theory became deeper and deeper with the discoveries of other types of electromagnetic radiation. X-rays discovered in 1898 by Roentgen, were shown in 1906 to be electromagnetic waves of wave length of order a few angstroms ($\sim 10^{-10}$m), much smaller than that of ordinary light waves. Our knowledge of em waves of various wavelength has continuously accumulated since then. Figures 9.10 and 9.11 represent the range of the wavelengths studied and the modern terminology used for various sections of the spectrum. The visible part of the spectrum, the one to which the human retina is sensitive, as is seen in Fig. 9.10, is a very tiny window in the whole spectrum of radiation. However, the whole spectrum is available to us for study and for use with different kinds of detectors that can detect various regions of the spectrum. Table 9.1 describes em waves used in telecommunication.

We know from our daily experience that some materials are transparent to light while others are not, implying of course that light radiation in travelling through any medium is absorbed in different amounts by different media. Not only that, there exist common place materials like tinted glass that allow light of some colours. We learn that when light travels through a material medium, the opacity is dependent both on the material and the wavelength of the radiation. Why is this so? Basically, light (or any electromagnetic wave) is absorbed by electric charges; the electric field of the electromagnetic wave can set the charge (which may be an electron, or an ion, or a dipolar molecule, or even a neutral atom or molecule) into forced oscillation. This transfers electromagnetic energy to the oscillator. Depending on the

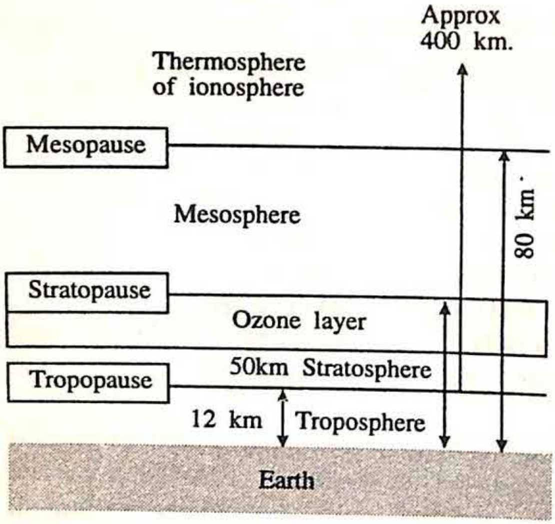


Figure 9.12: Various layers of our atmosphere. The actual boundaries are not sharp and the various distances given are not precise.

nature of the oscillator, i.e. the medium different wavelengths are absorbed differently. Thus for example, the human body which is opaque to visible light shows a different behaviour to X-rays - the tissues are quite transparent but the bones are relatively opaque - a feature fruitfully exploited in medical diagnosis. We will not discuss this general subject of interaction of radiation with media of various types any further but in the next section, will discuss a special medium, namely our atmosphere and its behaviour towards various kinds of radiation.

9.6 Electromagnetic radiation and the earth's atmosphere

The earth's atmosphere is a fascinating laboratory for the study of interaction of em radiation (both man - made and extraterrestrial) - with its various constituents. Although the background that we have achieved so far is not enough to completely understand the various phenomena involved, it is certainly worthwhile knowing a little about them - for reasons no other than that they involve us immediately.

As you know, the air thins out gradually as we go up - something that you can calculate using elementary physics. The atmosphere thus has no sharp boundary above, though it has several layers of regions with different properties, and names summarized in Fig. 9.12. Except for the layer in the upper atmosphere, called ionosphere, which is composed partly of electrons and +ve ions, the rest of the atmosphere is composed mostly of neutral molecules. The constitution of the different layers in the atmosphere is not the same. Water vapour, for example, is concentrated in the lowest layer whereas the ozone in the atmosphere is confined to the ozone layer, some 50-80 km above the ground.

The atmosphere is transparent to visible radiation since we can see the sun and the stars through it clearly. However, most infrared radiation is not allowed to pass through, i.e. the atmosphere absorbs it. Now, the energy from sunlight obviously heats the earth, which like any other hot body starts emitting radiation. However the earth is much cooler than the sun so that, according to Planck's law, its radiation is mostly in the infrared region unlike solar radiation. This radiation from the earth, however, is unable to cross the lower atmosphere, which reflects it right back. The earth's atmosphere is thus richer in infrared radiation which is sometimes called 'heat radiation' since most materials absorb them readily 'heating' themselves up in the process. Low lying clouds also prevent infrared radiation from passing through and thus serve to keep the earth's surface warm at night. This entire phenomenon is called the 'Greenhouse effect' (Fig. 9.13).

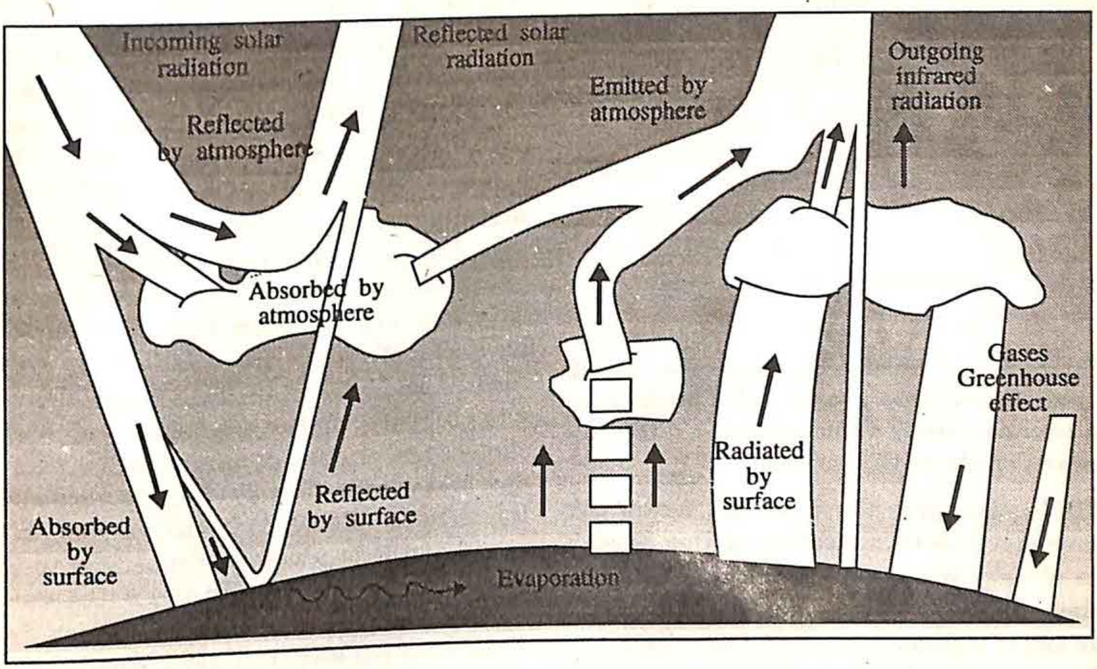


Figure 9.13: The balance of radiation: incoming, absorbed, reflected, radiated, emitted, etc. This describes the earth's energy balance including the effects of the atmospheric system.

The components of solar and other extra terrestrial sources in the ultraviolet (uv) and lower wavelength domains are dangerous - they cause genetic damage to living cells. The ozone layer blocks the passage of uv radiation and effectively protects us from the harmful portions of solar radiation. Practically all radiation of wavelength less than 3×10^{-7} m is absorbed by the ozone layer.

The behaviour of waves of wavelength 10^{-3} m and higher - broadly classified as radio waves - in their propagation through the atmosphere becomes relevant in all modern forms of communication: radio, television, microwaves etc. The earliest form of radio communication used electromagnetic waves of the type experimented upon by Hertz and Marconi - waves having wavelengths of 20 m or more. The atmosphere is more or less transparent to waves in this range - termed as AM (for amplitude modulated) band in modern terminology (Fig. 9.11). The topmost layer of the atmosphere, the ionosphere, however does not allow waves in this band to penetrate - they are reflected back. A signal emitted by an antenna from a certain point can thus be received at another point on the surface in two possible ways. The wave can travel directly following the surface of the earth (known as the 'ground wave') or can reach after being bounced back from the ionosphere (known as the sky wave). The groundwave attenuation increases with frequency, so that transmission via the ground wave is in practice possible for frequencies upto about 1500

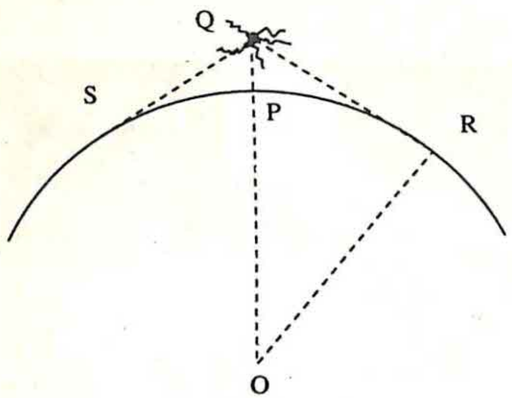


Figure 9.14: Line of sight transmission. Signals broadcast from atop a tower PQ of height h can be received directly within a distance $\sqrt{2Rh}$ on a receiver on the earth's surface.

kHz (wavelengths above 200 m). Below this wavelength, communication in this band is via the sky wave only. In common day language, these two regions of the AM band are normally referred to as medium wave and short wave bands.

Beyond a certain frequency - above 40MHz - the ionosphere bends any incident em radiation but does not reflect it back towards the earth. Since television signals have frequencies in the 100-200MHz range, transmission via the sky wave is impossible, Reception is possible only if the receiver antenna directly intercepts the signals. Thus if the broadcast is made from a height h above the ground (Fig. 9.14), no reception by direct signals is possible beyond points S and R in the figure.

The distance upto which signals can be received, PR is easily calculated in terms of h and the radius of the earth R. Considering the right angled triangle OQR, we have

$$\begin{aligned}(OQ)^2 &= (OR)^2 + (QR)^2;\\ OQ &= R + h, QR \simeq PR = d\\ OR &= R\end{aligned}$$

and hence

$$(R+h)^2 = R^2 + d^2$$

$$\text{or } d^2 = h^2 + 2hR$$

To get larger coverage, television broadcasts are thus made from tall antennas (i.e. large h), which is a familiar landmark in many cities. Fig. 9.15 shows, schematically, what happens in the atmosphere to radiowaves with different frequencies.

Radio-waves with frequencies higher than television signals are the microwaves- they have wavelengths of the order of a few millimeters. We know from day-to-day experience that a beam of light keeps along a straight path unlike a sound wave which spreads and also bends around the corners of any obstacles in its way. This phenomenon is due to the difference in the wavelength of sound and light waves in comparison to the size of the obstacles. Briefly, if the wavelength is small compared to the obstacle size, bending of the wave is negligible and it travels approximately in a straight line (ray or beam). But if the wavelength is comparable to or larger than that of the obstacle, the wave bends or diffracts a lot, as for example with sound waves in a room. Because of their relatively smaller wavelengths compared to radiowaves, microwaves are much better suited if one wants to 'beam' signals in a particular direction. This possibility was first successfully applied in a 'radar'. This device beams microwaves towards a distant object and receives the signal reflected back by the object. From a measurement of the time delay in receiving the signal back, the distance of the object is easily calculated since both the outgoing and incoming waves travel with the known velocity of light. Since the waves can be beamed in definite directions, the exact location of distant objects thus becomes known.

In more recent times, microwaves have revolutionalized telecommunications. Using artificial satellites, it has become possible to

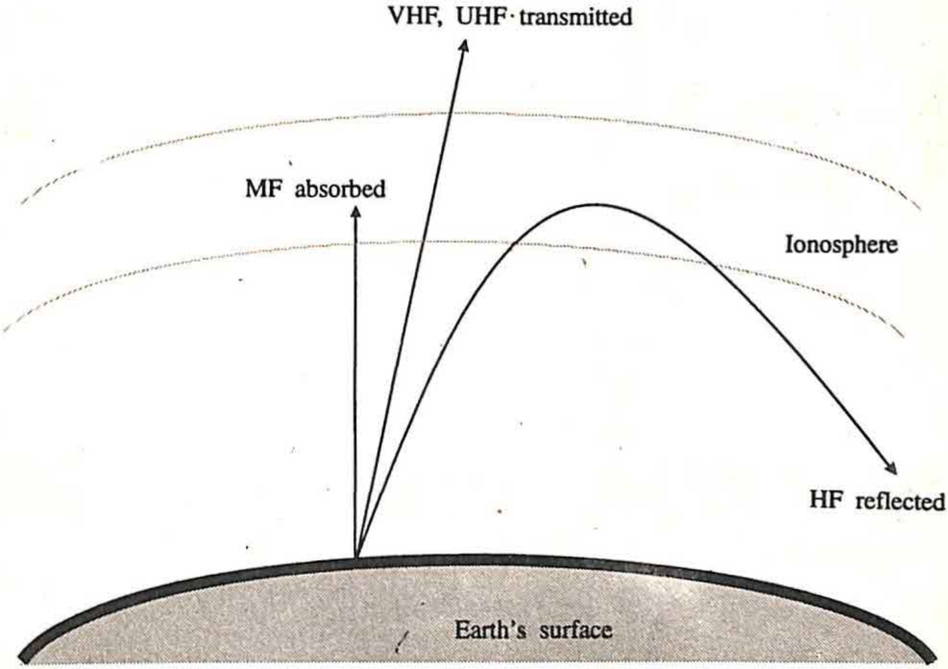


Figure 9.15: Effect of the ionosphere on electromagnetic waves (radio waves) in various frequency bands.

transmit signals from one point on the earth to practically any other point on the earth. The signals from the broadcasting station are beamed towards an artificial earth satellite, which in turn broadcasts it back to earth. Since the satellite is high above, it can send back the signals to a large part of the earth's surface. Of course, the satellite moves in its own orbit whereas the earth spins around its own axis. For a satellite to be useful for transmitting signals to definite regions, its orbit must be such that relative to the earth's surface it must appear fixed. Such a satellite must have a period of 24 hours in an orbit in the equatorial plane. The distance of such satellites is easily calculated to be some 36,000 km above the earth. Such satellites - called geostationary satellites - have been put up by many countries, including India, and lie parked densely in the equatorial plane at that height.

★ In section 9.4 we have seen that efficient radiation and propagation of electro magnetic waves is possible in the radiofrequency range of the electromagnetic spectrum. Here we discuss the basic principles of wireless radio communications - mainly the principle of amplitude modulation and demodulation.

The simplest scheme of wireless communication would be to convert the speech or music to be transmitted into electric signals using *a microphone*, boost up the power of the signal using *amplifiers* and radiate the signal in space with the aid of an *antenna*. This would constitute the *transmitter*. At the receiver end, one could have a pick-up antenna feeding the

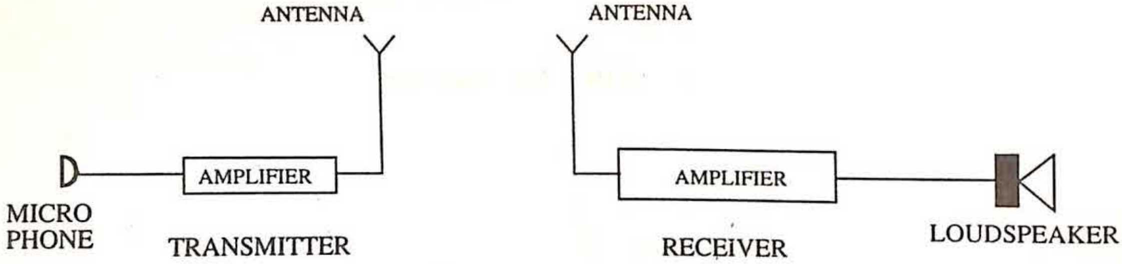


Figure 9.16:

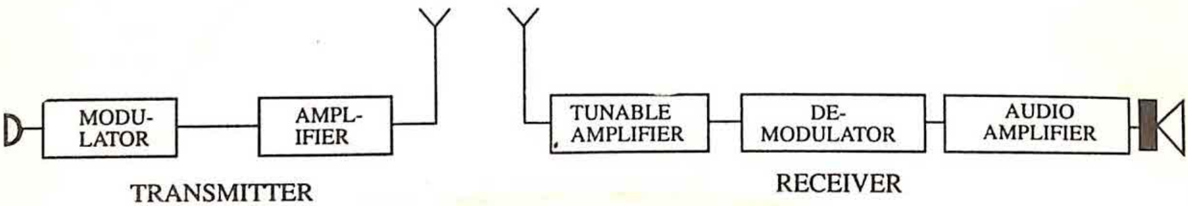


Figure 9.17:

speech or music signal to an amplifier and a loudspeaker (see Fig. 9.16).

However, this scheme suffers from the following drawbacks: (1) em waves in the frequency range of 20Hz - 20kHz (audio-frequency range) cannot be efficiently radiated and do not propagate well in space. (2) simultaneous transmission of different signals by different transmitters would lead to confusion at the receiver.

The way out of these difficulties is to devise methods to convert or translate the audio signals to the radio-frequency range before transmission and recover the audio-frequency signals back at the receiver. Different transmitting stations can then be allotted slots in the radio-frequency range and a single receiver can then tune in to these transmitters without confusion. The frequency range 500 kHz to 20 MHz is reserved for amplitude modulated broadcast, which is the range covered by most three band transistor radios. The process of frequency translation at the transmitter is called *modulation* and that of recovering the audio signal at the receiver is called *demodulation*. A simplified block diagram of such system is shown in Fig. 9.17.

Modulation: A radiofrequency signal of angular frequency ω_c can be represented as

$$v_c = V_{cm} \sin(\omega_c t + \phi)$$

The basic principle of modulation is to change some characteristic of this wave (called the carrier wave) in according with the audio-frequency signal (called the modulating signal). Accordingly, the methods of modulation fall into two classes: amplitude modulation and frequency modulation.

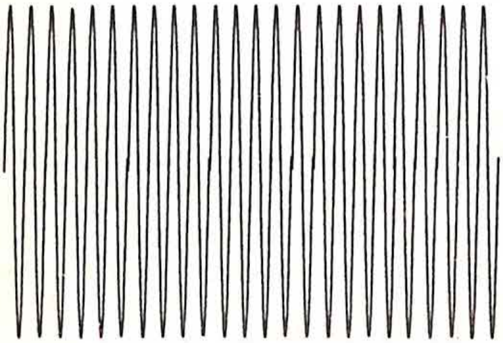
In an amplitude modulated signal, the amplitude of the carrier wave, V_{cm}, varies in accordance with the modulating signal. For simplicity in analysis, we shall take the modulating signal to be a purely sinusoidal signal with angular frequency ω_m : $v_m = V_{mm} \sin \omega_m t$. If $\phi = 0$ at $t = 0$, then we can write the modulated signal as

$$\begin{aligned} v_{mod} &= (V_{cm} + V_{mm} \sin \omega_m t) \sin \omega_c t \\ &= V_{cm}(1 + m \sin \omega_m t) \sin \omega_c t \end{aligned}$$

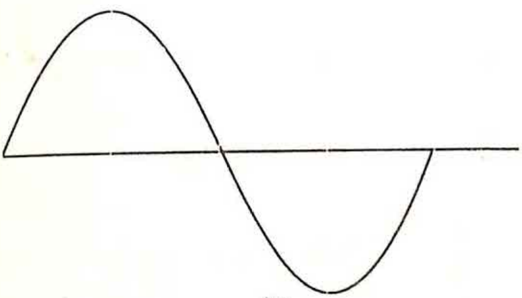
Where $m = (V_{mm}/V_{cm})$ measures the degrees of modulation. Notice that the amplitude modulated signal involves the *product*

$\sin\omega_m t \sin\omega_c t$. This is the signal which should be generated in the amplitude modulator. Figure 9.18 shows the carrier wave, the modulating wave and the modulated signal. The expression for V_{mod} gives us a clue for devising a method for amplitude modulation. We have seen that some devices have nonlinear voltage - current relationships. A device or a circuit having square law characteristics $i = av^2$(where i is the output current and v is the input voltage) can be used with $v = V_{cm}\ \sin\omega_c t + V_{mm}\sin\omega_m t$ as the input voltage, so that

$$\begin{aligned} i &= a(V_{cm}\sin\omega_c t + V_{mm}\sin\omega_m t)^2 \\ &= a(V_{cm}^2\sin^2\omega_c t + V_{mm}^2\sin^2\omega_m t \\ &\quad +2V_{cm}V_{mm}\sin\omega_c t\ \sin\omega_m t) \end{aligned}$$



(a)

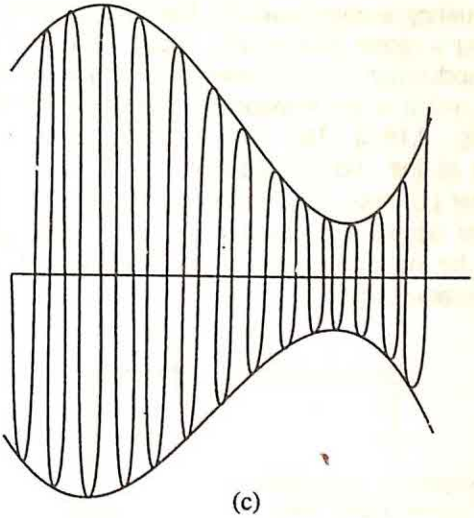


(b)

Figure 9.18: (a) The carrier wave and (b) the modulating wave.

Using the identity

$$\sin x\ \sin y = \frac{1}{2}\cos(x+y) + \frac{1}{2}\cos(x-y),$$



(c)

Figure 9.18: (c) AM wave with $m = 0.77$.

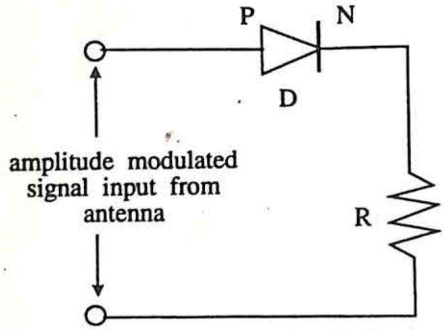
we have,

$$\begin{aligned} i = a\Bigg[&\frac{V_{cm}^2}{2}\cos 2\omega_c t + \frac{V_{cm}^2}{2} + \frac{V_{mm}^2}{2}\cos 2\omega_m t \\ +\ &\frac{V_{mm}^2}{2} + V_{cm}\ V_{mm}\cos(\omega_c + \omega_m)t \\ +\ &V_{cm}V_{mm}\cos(\omega_c - \omega_m)t)\Bigg]. \end{aligned}$$

We have the frequencies $2\omega_c, 2\omega_m$ (which are the second harmonics of the carrier and modulating signal respectively), $\omega_c + \omega_m$ and $\omega_c - \omega_m$ at the output of the modulator! The generation of harmonics is characteristic of all systems exhibiting nonlinear response. Recall that ω_c is in the radio-frequency range whereas ω_m is in the audio-frequency range. If $f_c = (\omega_c/2\pi)$ = 1 MHz, $f_m = (\omega_m/2\pi) =$ 1kHz, then we shall have signals with frequencies, 2 MHz, 2 kHz, 999kHz and 1001 kHz at the output of the square law modulator. The signal with frequency 2 kHz is not useful for transmission, and the signal with unwanted signals are filtered out using a parallel resonant circuit having the centre frequency of 1 MHz. The Q value of this resonant circuit is chosen so that the signals with frequencies 999kHz and 1001 kHz pass through to the amplifier without much attenuation.

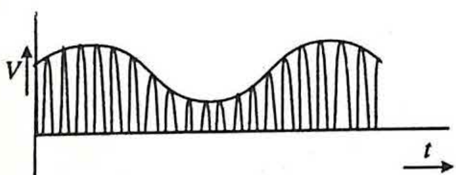
Demodulation: At the receiver end, the amplitude modulated signal of the desired carrier

frequency is demodulated: This is usually done using a diode and an RC circuit. The aim of demodulation is to recover the modulating signal, which is the envelope of the signal shown in Fig. 9.18(c). The diode D conducts only as long as the end labeled P in Fig. 9.19 is at a higher potential with respect to the end labeled N (for details see Chapter 14), i.e., it conducts only for the positive half-cycles of the amplitude modulated signal.



Action of a diode

(a)



(b)

Figure 9.19:

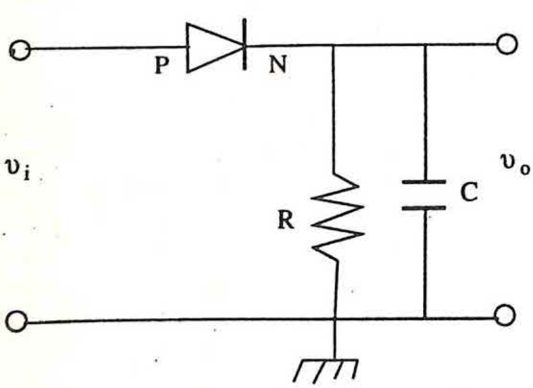
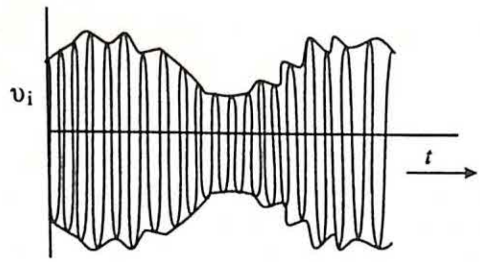
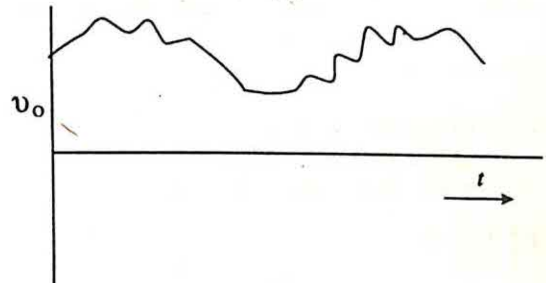


Figure 9.20:



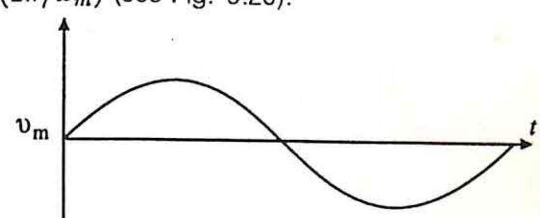
(a)



(b)

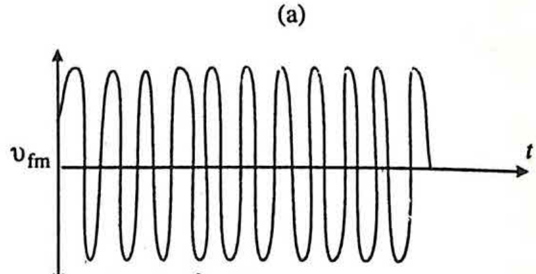
Figure 9.21: (a) The amplitude modulated input. (b) The demodulated output signal v_0. The ripple on the waveform is shown exaggerated.

To follow the envelope of the waveform shown in Fig. 9.19(b), a capacitor C is put in parallel with the resistor R such that $(2\pi/\omega_c) < RC < (2\pi/\omega_m)$ (see Fig. 9.20).



Modulating Signal

(a)



Frequency modulated signal

(b)

Figure 9.22:

As shown in Fig. 9.21(b), the capacitor C

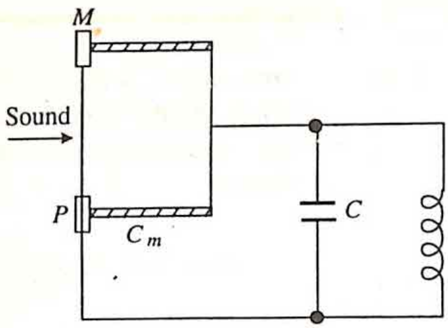


Figure 9.23:

charges to the positive peak of the amplitude-modulated waveform and then starts discharging Since the time period of the radio-frequency signal is much smaller than the RC time constant, the voltage across the RC circuit goes down by a very small amount before the next peak of the radio-frequency signal starts charging the capacitor again. In this way, the voltage across the RC circuit roughly follows the envelop of the waveform shown in Fig. 9.19(b) and recovers the modulating signal. This signal is then amplified and fed to a loudspeaker.

In frequency modulation, the instantaneous frequency of the carrier wave is changed in accordance with the amplitude of the modulating signs (see Fig. 9.22). We shall not go into the details of methods of frequency modulation and demodulation, but restrict ourselves to a simple frequency modulator which is sometimes used in wireless microphone public address systems.

Consider the system shown in Fig. 9.23: a capacitor microphone M is placed in parallel, with L and C. The resonant frequency of this parallel tuned circuit is $f_R = 1/(2\pi\sqrt{L(C+C_M)})$. The plate P of the capacitor microphone responds to variations in pressure due to sound waves. Thus C_M changes with the amplitude of the sound signal. This parallel tuned system is used as the frequency determining part of an a.c. generator (called an *oscillator*) to obtain the frequency modulated signal.

The sound signal in television transmissions is frequency modulated. ★

Summary

1. *Displacement current*: Maxwell noticed that Ampere's circuital law is inconsistent namely, makes nonunique predictions for the magnetic field in situations where the electric current changes with time. He showed that consistency required an additonal source of magnetic field; this is called the *displacement current.* Eq. (5.29) is now modifed, and states in its simplified form:

   If $\mathbf{B}$ is always directed along the tangent to the perimeter of a closed curve, *and,* has a constant magnitude B all along the perimeter,

   $B \times$ (perimeter of the closed circuit)

   $= \mu_0 \times$ (total current I passing through *any* surface with the closed circuit as boundary).

   The total current I is the sum of the conduction current I_C carried by charges moving in conductors, *and* the displacement current I_d which is ϵ_0 times the rate of change of the electric flux through the surface mentioned above.

   $$\begin{aligned} I &= I_c + I_d \\ &= I_c + \epsilon_0 (d\Phi_E/dt) \end{aligned}$$

   Roughly, the existence of the displacement current means that time dependent electric fields produce magnetic fields. This is to be compared with Faraday's law of electromagnetic induction, which implies that time dependent magnetic fields produce electric fields. Thus the laws of electricity and magnetism are quite symmetric.

2. *Electromagnetic waves*: The displacement current as a source of magnetic field, and the resulting symmetry of the laws of electricity and magnetism led Maxwell to the conclusion that electromagnetic waves must exist. Electric and magnetic fields oscillate sinusoidally with space and time in an electromagnetic wave, giving rise to and sustaining each other. The oscillating electric and magneit fields $\mathbf{E}$, $\mathbf{B}$ are perpendicular to each other, and to the direction of propagation of the electromagnetic wave. If the wave (of wavelength $\lambda = (2\pi/k)$) propagates along the z direction with frequency ν (angular frequency $\omega = 2\pi\nu$) then

   $$E = E_x(t) = E_{x_0} \sin(kz - \omega t) = E_{x_0} \sin\left(\frac{2\pi}{\lambda} z - 2\pi\nu t\right)$$

   and

   $$B = B_y(t) = B_{y_0} \sin(kz - \omega t) = B_{y_0} \sin\left(\frac{2\pi}{\lambda} z - 2\pi\nu t\right)$$

   The quantities ν and λ are related to the velocity c of the wave by the usual relation

   $$\nu\lambda = c$$

   or $\omega = ck$

The velocity c of electromagnetic waves in vacuum or free space is related to μ_0 and ϵ_0 (the magnetic permeability and dielectric permittivity constants) as follows

$$c = 1/\sqrt{\mu_0\epsilon_0}$$

This relation can be obtained by dimensional arguments.

Light is an electromagnetic wave; c is therefore also the velocity of light. Electromagnetic waves other than light also have the same velocity c in free space.

The velocity of light, or of electromagnetic waves in a material medium is given by

$$c = 1/\sqrt{\mu\epsilon}$$

where μ is the permeability of the medium and ϵ is its dielectric constant.

3. *Production and detection of electromagnetic waves*: Accelerated electric charges produce electromagentic waves. An electric charge oscillating harmonically with frequency ν, provides electromagnetic waves of the same frequency ν. Such an *electric dipole* is a basic source of electromagnetic waves. It is an efficient source if the size of the oscillating dipole is of the same order as the wavelength of the em wave (Generally the sources are smaller, often much smaller).

   Radio waves (em waves with wavelength in the range a few mm to a few hundred metres) have as their sources electric dipoles (eg. two pieces of conducting wire) connected to an ac circuit of appropirate resonant frequency.

   Radio waves are detected by as electric dipole antenna or a (magnetic) loop antenna. In the former, the electric field of the incident em wave causes the electrons to oscillate, and sets up an ac signal which is detected. In a loop antenna, the oscillating magnetic field of the em wave sets up an ac emf which is detected.

   Electromagnetic waves with wavelength of the order of a few metres were first produced and detected in the laboratory by Hertz in 1887. He showed that these have exactly the same physical properties as light. Electromagnetic waves in this wavelength region are widely used now in telecommunication.

4. *The electromagnetic spectrum*: The spectrum of electromagnetic waves stretches in principle over an infinite range of wavelengths. Different regions are known by different name: γ rays, X-rays, ultraviolet rays, visible light, infrared rays, microwaves and radio waves in order of increasing wavelength from 10^{-2} Åor 10^{-12} m to 10^6m.

   They interact with matter via their electric and magnetic fields which set in oscillation charges present in all matter. The detailed interaction and so the mechanism of absorption, scattering etc. depend on the wavelength of the em wave, and the nature of the atoms and molecules in the antena medium.

5. *The atmosphere*: A medium of particular importance is the atmosphere. It is transparent to visible light. The ozone layer absorbs harmful ultraviolet light. Infrared waves both from the sun and re-emitted from the earth are partially absorbed by

atmospheric gases such as carbon dioxide, and by clouds. The energy balance near the earth's surface is affected by man's activities, and can lead to changes such as global warning or ozone depletion.

The propagation of radio waves emitted from the earth depends on the wavelength of the waves. Medium frequency (MF) waves (300 kHz - 3 MHz) are largely absorbed, the high frequency (HF) waves (3-30 MHz) are reflected back by the ionosphere, and shorter wavelength waves go straight through it. The last (VHF and UHF waves in the range 30 MHz to 3 GHz) are transmitted from the one place to another either by direct line of sight using tall towers, or by beaming to artificial satellites and rebroadcasting from there.

Exercises

9.1 What physical quantity is the same for X-rays of wavelength 10^{-10}m, red light of wavelength 6800Åand radiowaves of wavelength 500m?

9.2 A plane electromagnetic wave travels in vacuum along z-direction. What can you say about the directions of its electric and magnetic field vector? If the frequency of the wave is 30 MHz. What is its wavelength?

9.3 A radio can tune to any station in the 7.5 MHz to 12 MHz band. What is the corresponding wavelength band?

9.4 A parallel-plate capacitor is being charged by an external source. Show that the sum of conduction current and displacement current has the same value everywhere in the circuit.

9.5 A charged particle oscillates about its mean equillibrium position with a frequency of 10^9Hz. What is the frequency of the electromagnetic waves produced by the oscillator?

9.6 Give a simple plausibility argument to suggest that an accelerated charge must emit electromagnetic radiation.

9.7 A TV tower has height of 100m. How much population is covered by the TV broadcast if the average population density around the tower is 1000 km^{-2} (radius of the earth = 6.37 x 10^6m).

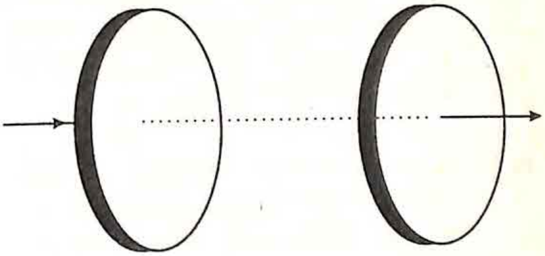
Additional Exercises

9.8 Figure shows a capacitor made of two circular plates each of radius 12 cm, and seperated by 5.0 mm. The capacitor is being charged by an external source (not shown in the figure). The charging current is constant and equal to 0.15A.

(a) Calculate the capacitance and the rate of change of potential difference between the plates.

(b) Obtain the displacement current across the plates.

(c) Is Kirchhoff's first rule valid at each plate of the capacitor? Explain.



9.9 (a) Refer to the figure for exercise 9.8. Use Ampere's law (modified to include displacement current as given in the text) and the symmetry in the problem to calculate magnetic field between the plates at a point (i) on the axis, (ii) 6.5 cm from the axis, and (iii) 15 cm from the axis.

(b) At what distance from the axis is the magnetic field due to displacement current maximum? Obtain the maximum value of the field.

9.10 (a) Use the Biot-Savart law to determine the magnetic field due to the conduction current outside the plates (refer to the figure for exercise 9.8) at points 6.5 cm, 12 cm and 15 cm from the wire. Do the answers match with those in exercise 9.9? Explain.
(b) If the conducting wire has a radius of 1.0 mm, what is the maximum value of magnetic field due to the conduction current? [When you compare the answers to exercise 9.9 and 9.10, you will appreciate why it is not easy to notice magnetic field due to the displacement current.].
(c) Suppose the thin wire in figure for 9.8 is replaced by rods each of radius 12 cm (i.e. we now have two long cylindrical rods separated by a small gap). Will the magnetic field configurations for $r \geq R$ be identical for the regions between the plates and outside the plates?

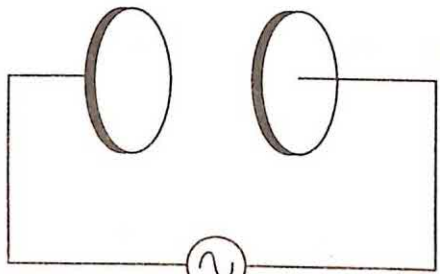
9.11 A parallel plate capacitor of area 50 cm^2 and plate separation 3.0 mm is charged initially to 80 μC. Due to a radioactive source nearby, the medium between the plates gets slightly conducting and the plate loses charge initially at the rate of 1.5 x 10^{-8} C s^{-1}. What is the magnitude and direction of displacement current? What is the magnetic field between the plates?

9.12 A parallel plate capacitor made of circular plates each of radius R = 6.0 cm has a capacitance C = 100 pF. The capacitor is connected to a 230 V a.c. supply with a (angular) frequency of 300 rad s^{-1}.

(a) What is the rms value of the conduction current?

(b) Is the conduction current equal to the displacement current?

(c) Deterrmine the amplitude of **B** at a point 3.0 cm from the axis between the plates.



9.13 The four maxwell's equations and the Lorentz force law (which together constitute the foundation of all of classical electromagnetism) are listed below:
(1) $\oint \mathbf{E}.\, ds = q/\epsilon_o$
(2) $\oint \mathbf{B}.\, ds = 0$
(3) $\oint \mathbf{E}.\, dl = -\frac{d}{dt}\int \mathbf{B}.\, ds$
(4) $\oint \mathbf{B}.\, dl = \mu_o I + \mu_o\epsilon_o\frac{d}{dt}\int \mathbf{E}.\, ds$

Lorentz force law: $\mathbf{F} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$

Answer the following questions regarding these equations:

(a) Give the name(s) associated with some of the four equations above.

(b) Which equations above contain sources of **E** and **B** and which do not? What do the equations reduce to in a source-free region?

(c) Write down Maxwell's equations for steady (i.e. time independent) electric and magnetic fields.

(d) If magnetic monopole's existed, which of the equations would be modified? Suggest how they might be modified.

(e) Which of the four equations shows that magnetic field lines cannot start from a point nor end at a point?

(f) Which of the four equations shows that electrostatic field lines cannot form closed loops?

(g) The equations listed above refer to integrals of **E** and **B** over loops/surfaces. Can we write down equations for **E** and **B** for each point in space?

(h) Are the equations listed above true for different types of media: dielectrics, conductors, plasmas etc.?

(i) Are the equations true for arbitrarily high and low values of **E**, **B**, q, I?

9.14 Consider a plane wave-front of electromagnetic fields travelling with a speed c in the right (say + z) direction; it is given that **B** and **E** are transverse to each other and uniform throughout the left of the wavefront and zero on the right of the wavefront. [This is a contrived, but not incorrect, configuration chosen for simplicity. In the usual monochromatic plane wave, **E** and **B** vary sinusoidally in space and time].

(a) Use Faraday's law to show that $E = cB$

(b) Use Ampere's law (with displacement current included) to show that $c = 1/\sqrt{\mu_o \epsilon_o}$.

9.15 (a) Let us take it as given that Maxwell's equations predict the existence of electromagnetic waves in vacuum with a constant speed $c = 1/\sqrt{\mu_o \epsilon_o} = 3 \times 10^8 \text{ms}^{-1}$. Do you find something strange in the appearance of a constant in the theory that has dimensions of speed? Explain.

(b) The hypothesis of ether (a medium permeating all of space, and endowed with very special properties) is a natural idea based on Maxwell's equations (though we know now that the hypothesis is incorrect). Explain why the ether is such a natural idea?

(c) Are Maxwell's equations modified by Einstein's special relativity? Are they valid in the domain of quantum theory?

9.16 In a plane em wave, the electric field oscillates sinusoidally at a frequency of 2.0 x 10^{10}Hz and amplitude 48Vm^{-1}.

(a) What is the wavelength of a wave?

(b) What is the amplitude of the oscillating magnetic field?

(c) Show that the average energy density of the **E** field equals the average energy density of the **B** field. [$c = 3 \times 10^8$ ms^{-1}].

(**Note:** The relation connecting amplitudes of **E** and **B** fields for a plane monochromatic wave is the same as given in Eq.(9.13).

9.17 Answer the following questions carefully:

(a) Magnetic field lines can never emanate from a point nor end on a point. Yet the field lines outside a bar magnet do seem to start from the North pole and end on the South pole. Does the second fact contradict the first? Explain.

(b) If you find closed loops of **B** in a region in space, does it necessarily mean that actual charges are flowing across the area bounded by the loops?

(c) A closed loop of **B** is produced by a changing electric field. Does it necessarily mean that **E** and d**E**/dt are non-zero at all points on the loop and in the area enclosed by the loops?

(d) Why is it that induced electric fields due to changing magnetic flux are more readily observable than the induced magnetic fields due to changing electric fields?

(e) A variable-frequency ac source is connected to a capacitor. Will the displacement current increase or decrease with increase in frequency?

(f) Electromagnetic waves in a cavity with conducting walls can exist only in certain modes (i.e. they cannot exist, for example, with any argbitrary wavelength). Suggest a simple reason why this should be so.

9.18 The terminology for different parts of the electromagnetic spectrum is given in the text. Use the formula $E = h\nu$ (for energy of a quantum of radiation:photon) and obtain the photon energy in units of eV for different parts of the em spectrum. In what way are the different scales of photon energies that you obtain related to the sources of electromagnetic radiation?

9.19 Use the formula $\lambda_m T = 0.29$ cm K to obtain the characteristic temperature ranges for different parts of the em spectrum. What do the numbers that you obtain tell you? [See the answer for an explanation of the formula].

9.20 Given below are some famous numbers associated with electromagnetic radiation in different contexts in physics. State the part of the em spectrum to which each belongs.

(i) 21cm (wavelength emitted by atomic hydrogen in interstellar space).

(ii) 1057 MHz [frequency of radiation arising from two close energy levels in hydrogen; known as Lamb

shift].
(iii) 2.7 K [temperaturre associated with the isotropic radiation filling all space- thought to be a relic of the 'big-bang' origin of the universe].
(iv) 5890 Å- 5896 Å[double lines of sodium]
(v) 14.4 keV [energy of a particular transition in ^{57}Fe nucleus associated with a famous high resolution spectroscopic method (Mössbauer spectroscopy].

9.21 Answer the following questions:

(a) Long distance radio broabcasts use short-wave bands. Why?

(b) It is necessary to use satellites for long distance TV transmission. Why?

(c) Optical and radiotelescopes are built on the ground but X-ray astronomy is possible only from satellites orbiting the earth. Why?

(d) The small ozone layer on top of the stratosphere is crucial for human survival. Why?

(e) If the earth did not have an atmosphere, would its average surface temperature be higher or lower than what it is now?

(f) Some scientists have predicted that a global nuclear war on the earth would be followed by a severe 'nuclear winter' with a devastating effect on life on earth. What might be the basis of this prediction?

9.22 Check the dimensional consistency of Maxwell's equations and the Lorentz force law given in exercise 9.13.

Hertz, Heinrich Rudolf (1857-1894) German physicist who was the first to broadcast and receive radio waves. He produced electromagnetic waves, sent them through space, and measured their wavelength and velocity. He showed that the nature of their vibration, reflection and refraction was the same as that of light and heat waves, establishing their identity for the first time. He also pioneered research on discharge of electricity through gases, and discovered the photoelectric effect.

Bose, Jagadish Chandra (1858-1937) Indian physicist and plant physiologist. He produced, in 1895, ultrashort radio waves, and demonstrated their quasioptical properties. He invented highly sensitive instruments for the detection of minute responses by living organisms to external stimuli. He anticipated from his results the parallelism between animal and plant tissues noted by later biophysicists.

CHAPTER **10**

Wave Optics

10.1 Introduction

Simple and common observations such as the shape, size and sharpness of shadows - led to the idea of *rays* of light. Newton tried to understand travel of light in a straight line, reflection, and refraction assuming that it was made up of particles. However, starting from the seventeenth century, new kinds of behaviour of light were discovered experimentally. These could not be understood using simple particle ideas, but needed the assumption of *waves*. Later, in the nineteenth century, the study of electricity and magnetism led to Maxwell's idea of electromagnetic waves. These were experimentally verified by Hertz (Chapter 9 of this book). It was realised by Maxwell and Hertz that visible light is also made up of waves of this kind. The wavelength varies from 390 nm (3.9×10^{-7} m) for violet to 760 nm (7.6×10^{-7} m) for red.

The wave-like behaviour of light is the subject of this chapter. Some of the effects described are common to all kinds of waves including sound waves. For example, light falling on a small slit or hole spreads out

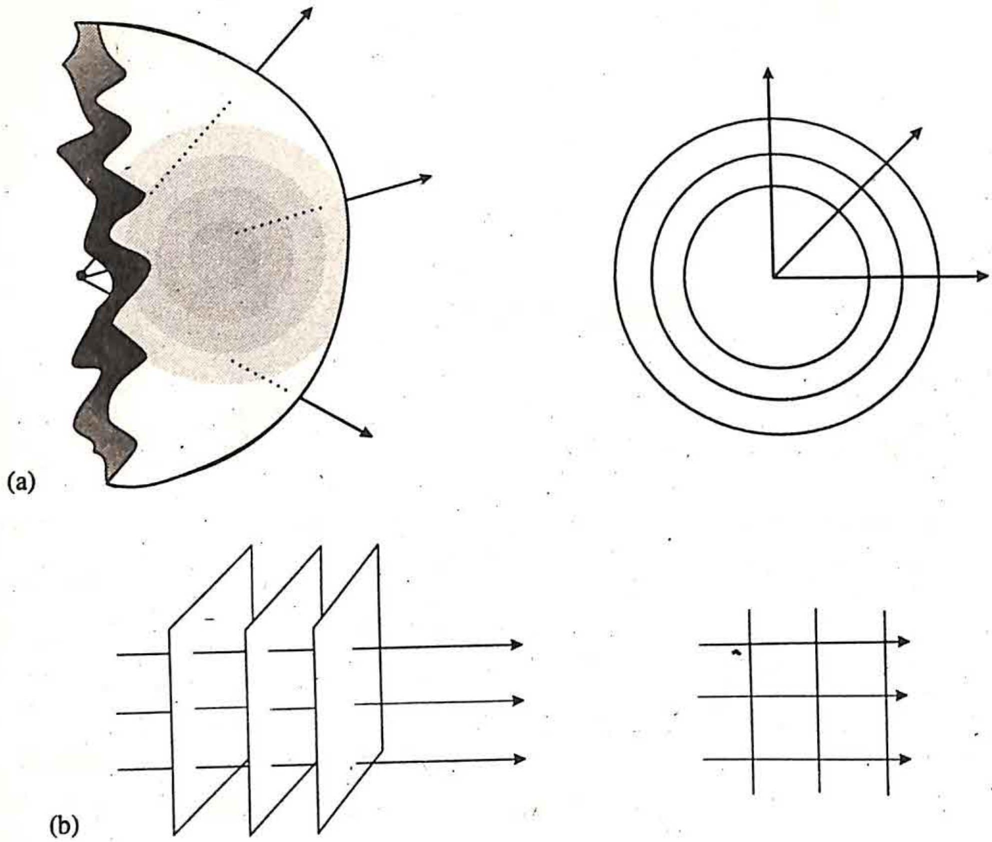


Figure 10.1: Wavefronts and the corresponding rays in two cases: (a) diverging spherical wave, (b) plane wave. The figure on the left shows a wave (e.g., light) in three dimensions. The figure on the right shows a wave in two dimensions (a water surface).

into an angle which becomes larger as the openings get narrower. This phenomenon is called *diffraction*. Another remarkable effect is *interference*, which you have already come across in Chapter 13 (Class XI text book). Waves from two different sources reaching the same point can combine to produce an intensity which may be less than the sum of the individual intensities (even zero!) or more than the sum. The *Doppler effect* is again common to all waves. The frequency observed is different from the frequency emitted, depending on the relative motion of the source and the observer.

There are also interesting effects which occur only for transverse waves. The behaviour of light in reflection, refraction, and transmission through crystals was inferred to depend on the direction of the electric field of the light wave in the plane perpendicular to the direction of travel. This property is called *polarisation* of light.

We will describe all these interesting wave optical effects in more detail later in this chapter. But we start by explaining the more familiar properties of light - propagation in straight lines, reflection and refraction - using waves. Christian Huygens of Holland, father of wave optics, was the first to do so in the late seventeenth century and

his ideas are described in the next section.

10.2 Rays and wavefronts

Consider a wave spreading out on the surface of water after a stone is thrown in. Every point on the surface oscillates. At any time, a photograph of the surface would show circular rings on which the disturbance is a maximum. Clearly, all points on such a circle are oscillating in phase because they are at the same distance from the source. Such a line connecting points which oscillate in phase is an example of a *wavefront*. *A wavefront is defined as a surface of constant phase*. The speed with which the wavefront moves outwards from the source is called the *phase velocity*. The energy of the wave moves in a direction perpendicular to the wavefront.

Fig. 10.1 shows light waves from a point source forming a spherical wavefront in three dimensional space. The energy travels outwards along straight lines emerging from the source, that is, radii of the spherical wavefront. These lines are the rays. Notice that when we measure the spacing between a pair of wavefronts along any ray, the result is a constant. This example illustrates two important general principles which we will use later:

(i) *Rays are perpendicular to wavefronts,*

(ii) *The time taken for light to travel from one wavefront to another is the same along any ray.*

If we look at a small portion of a spherical wave, far away from the source, then the wavefronts are like parallel planes. The rays are parallel lines perpendicular to the wavefronts. This is called a plane wave and is also sketched in Fig. 10.1.

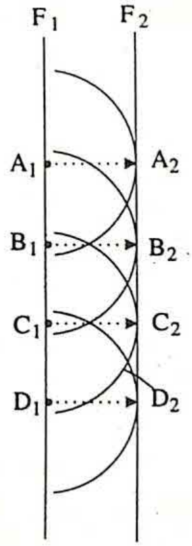


Figure 10.2: Huygens' geometrical construction for wave propagation. F_1 is a wavefront at some time, and F_2 at a time t later. The lines A_1A_2, B_1B_2 etc., are normal to both F_1 and F_2 and represent rays.

10.2.1 Huygens' construction

Huygens gave a beautiful geometrical description of wave propagation. This is illustrated in Fig. 10.2, in the simple case of a plane wave.

(a) At time $t = 0$, draw a surface F_1 which we call a *front*. It separates those parts of the medium which are undisturbed from those where the wave has already reached.

(b) Each point on F_1 acts like a new source and sends out a spherical wave of radius vt. This is called a secondary wave.

(c) After a time t, the disturbance would now have reached all points within the region covered by all these secondary waves. The boundary of this region is the new front F_2. Notice that F_2 is a

surface tangent to all the spheres. It is called the *forward envelope* of these secondary wavefronts.

(d) The secondary wave from the point A_1 on F_1 touches F_2 at A_2. The line connecting any point A_1 on F_1 to the corresponding point A_2 on F_2 is a ray. It is perpendicular to F_1 and F_2 and has length vt. This explains why rays are perpendicular to wavefronts and the time between two wavefronts is the same along any ray.

(e) The construction can be repeated starting with F_2 to get the next wavefront F_3 a time t later, and so on.

Huygens' construction can be understood physically for waves in a medium. Each oscillating particle acts as a "secondary" source, setting its neighbours into oscillation. The mathematical theory (not given here) shows that the geometrical construction works in the same way even for waves in vacuum!

★ The construction is more general than our simple example. The wavefronts can have any shape. The speed could be different at different places. In crystals, the speed is even different in different directions. Crystals of calcite (calcium carbonate, $CaCO_3$) show *double refraction* (two refracted rays as in Fig. 10.3). Huygens was able to explain the geometry of these rays with *two* wavefronts, one a sphere and the other a spheroid, present simultaneously. ★

10.2.2 Reflection and refraction

We can now understand the laws of reflection and refraction using the wave theory. Fig. 10.4 shows the incident and reflected wavefronts when a parallel beam of light falls on

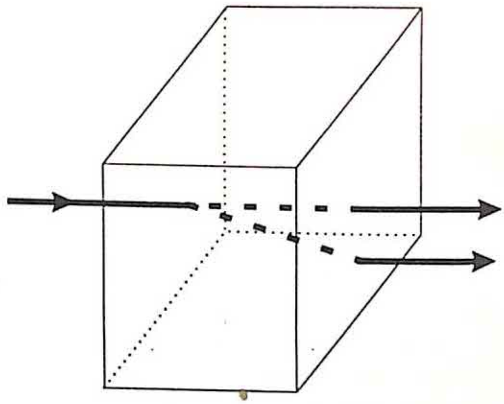


Figure 10.3: Double refraction exhibited by crystals like calcite. A single incident ray gives rise to two refracted rays.

a plane surface. One ray POQ is shown normal to both the reflected and incident wavefronts. The angle of incidence i and the angle of reflection r are defined as the angles made by the incident and reflected rays with the normal. As shown in Fig. 10.4, these are also the angles between the wavefront and the surface.

We now calculate the total time to go from one wavefront to another along the rays. From Fig. 10.4, we have

$$\begin{aligned}
&\text{Total time from P to Q}\\
&= \frac{PO}{v_1} + \frac{OQ}{v_1}\\
&= \frac{AO\sin i}{v_1} + \frac{OB\sin r}{v_1}\\
&= \frac{OA\sin i + (AB - OA)\sin r}{v_1}\\
&= \frac{AB\sin r + OA(\sin i - \sin r)}{v_1}. \qquad (10.1)
\end{aligned}$$

Different rays normal to the incident wavefront strike the surface at different points O and hence have different values of OA. Since the time should be the same for all the rays, the right side of Eq.(10.1) must actually be independent of OA. The condition for this

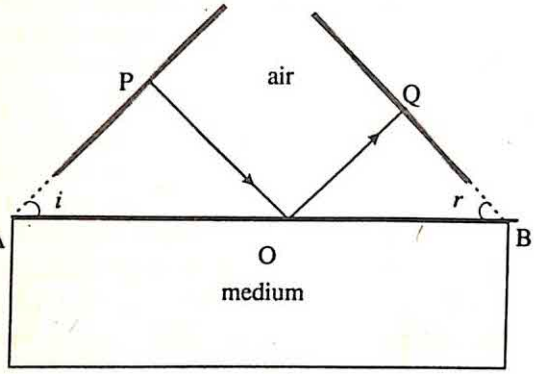


Figure 10.4: Wavefronts and corresponding rays for reflection at a place surface.

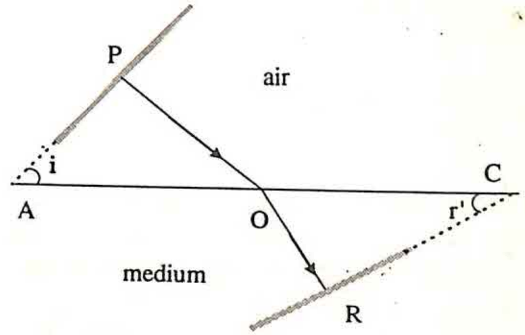


Figure 10.5: Wavefronts and corresponding rays for refraction by a plane surface separating two media.

to happen is that the coefficient of OA in Eq.(10.1) should be zero *i.e.*, $\sin i = \sin r$. We thus have the law of reflection

$$i = r. \tag{10.2}$$

We now come to refraction. Fig. 10.5 shows a plane surface separating medium 1 (speed of light v_1) from medium 2 (speed of light v_2). The incident and refracted wavefronts are shown making angles i and r' with the boundary. r' is called the angle of refraction. Rays perpendicular to these are also drawn. As before, let us calculate the time taken to travel between the two wavefronts along any ray.

Time taken from P to R $= \dfrac{PO}{v_1} + \dfrac{OR}{v_2}$

$$= \frac{OA \sin i}{v_1} + \frac{(AC - OA)\sin r'}{v_2}$$
$$= \frac{AC}{v_2}\sin r' + OA\left(\frac{\sin i}{v_1} - \frac{\sin r'}{v_2}\right) \tag{10.3}$$

This time should again be independent of which ray we consider. The coefficient of OA in Eq.(10.3) is therefore zero.

$$\frac{\sin i}{\sin r'} = \frac{v_1}{v_2} \tag{10.4}$$

Notice that Eq.(10.4) is Snell's law of refraction. The ratio of the phase velocity of light c in vacuum to its value v_1 in a medium is called the refractive index n_1 of the medium

$$n_1 = c/v_1 \tag{10.5}$$

For example, the refractive index of water for visible light is about 1.33. This means that the speed in water is $v_1 = c/n_1$ i.e., 0.75 times the speed in vacuum. (The refractive index of vacuum is clearly equal to 1). The measurement of the speed of light in water by Foucault (1850) confirmed this prediction of the wave theory. We have seen in Chapter 9 (see Eq. (9.8) and discussion there) that the velocity of light in a medium depends on the dielectric constant $K\{=(\epsilon/\epsilon_0)\}$ and relative magnetic permeability $K_m\{=(\mu/\mu_0)\}$. There is, therefore, a direct connection between these two and the refractive index n. Namely,

$$n = \sqrt{KK_m} \tag{10.6}$$

where the quantities K and K_m are measured at the frequency of light.

When light travels from medium 1 to medium 2, we measure the refractive index of medium 2 relative to medium 1, denoted by n_{12}

$$\frac{\sin i}{\sin r'} = n_{12} = \frac{v_1}{v_2} = \frac{(c/v_2)}{(c/v_1)} = \frac{n_2}{n_1} \qquad (10.7)$$

This equation expresses n_{12} in terms of n_1 and n_2, the refractive indices of the two media relative to vacuum. In the laboratory, medium 1 is usually air, with a refractive index of 1.0003, quite close to 1. For this reason, we can usually put n_1 in the denominator of Eq.(10.6) equal to 1, and treat the value of $\sin i/\sin r'$ measured in the laboratory as n_2, the refractive index of medium 2.

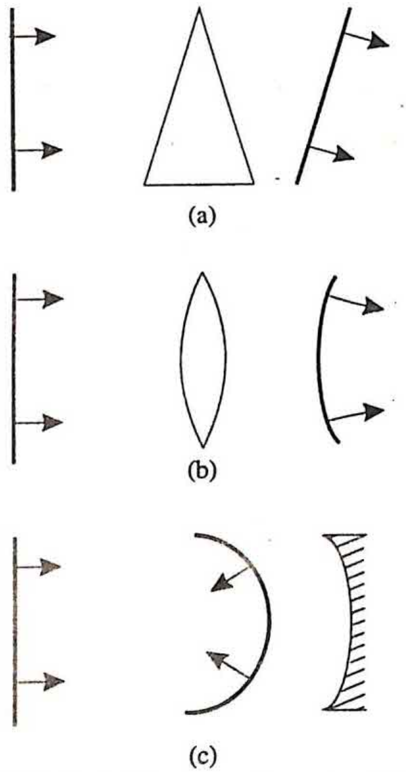


Figure 10.6: Action on a plane wavefront due to (a) a thin prism, (b) a convex lens, and (c) a concave mirror.

Let a source of light be at rest in one medium and the observer at rest in another medium. The time taken to travel between them is then fixed. Let two wavefronts separated by a whole cycle of phase (*i.e.*, 2π) be emitted from the source, separated by a time T. Their arrival at the observer is also separated by the same time interval T. *The frequency $\nu(= 1/T)$ therefore remains the same as light travels from one medium to another*, *i.e.*, $\nu_1 = \nu_2$. Because the velocities of light v_1 and v_2 are different, the wavelengths λ_1 and λ_2 are also different. Using the relation $v = \nu\lambda$,

$$\frac{v_1}{v_2} = \frac{\nu_1\lambda_1}{\nu_2\lambda_2} = \frac{\lambda_1}{\lambda_2}. \qquad (10.8)$$

The wavelength in a medium is directly proportional to the phase velocity and hence inversely proportional to the refractive index.

Once we have the laws of reflection and refraction, the behaviour of prisms, lenses, and mirrors can be understood. These topics are discussed in detail in the next chapter (Chapter 11). Here we just describe the behaviour of the wavefronts in these three cases (Fig. 10.6).

(i) Consider a plane wave passing through a thin prism. Clearly the portion of the incoming wavefront which travels through the greatest thickness of glass has been delayed the most, since light travels more slowly in glass. This explains the tilt in the emerging wavefront.

(ii) Similarly, the central part of an incident plane wave traverses the thickest portion of a convex lens and is delayed the most. The emerging wavefront has a depression at the centre. It is spherical and converges to a focus.

(iii) A concave mirror achieves the same effect as a convex lens. The centre of the wavefront has to travel a greater

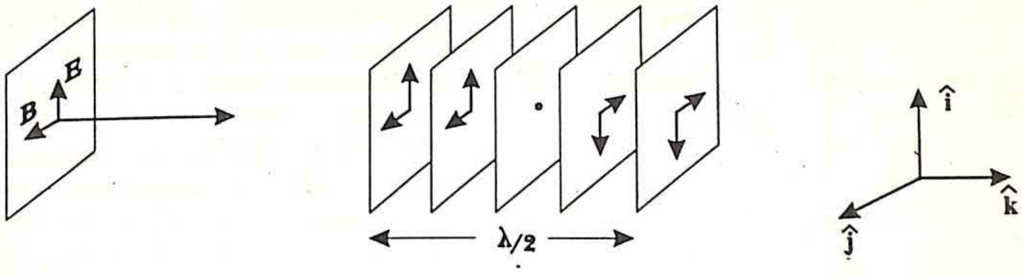


Figure 10.7: A plane electromagnetic wave propagating along z. The wavefronts are shown parallel to the $x-y$ plane. The directions of $\mathbf{E}$ and $\mathbf{B}$ are shown at the left and their space variation at the right. The long arrow along z is the direction of propagation.

distance before and after getting reflected, when compared to the edge. This again produces a converging spherical wavefront.

(iv) Concave lenses and convex mirrors can be understood from time delay arguments in a similar manner.

10.3 Basic properties of electromagnetic waves

10.3.1 A plane wave

We now recollect some basic properties of electromagnetic waves (Chapter 9). We describe the simplest case - a plane wave with a frequency $\nu = 1/T$, travelling along the z direction. We give below two formulae (Eq. (9.9)) which give the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ at a point (x, y, z) at a time t.

$$\mathbf{E}(x,y,z,t) = E_0\hat{\mathbf{i}}\sin\left(\frac{2\pi t}{T} - \frac{2\pi z}{\lambda} + \varphi_0\right) \quad (10.9a)$$

$$\mathbf{B}(x,y,z,t) = B_0\hat{\mathbf{j}}\sin\left(\frac{2\pi t}{T} - \frac{2\pi z}{\lambda} + \varphi_0\right) \quad (10.9b)$$

In these equations, E_0 and B_0 are the amplitudes (*i.e.*, maximum values) of the electric field and magnetic field. Unit vectors along the three coordinate directions are written as $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$. Note that the electric field is perpendicular to the magnetic field and both are perpendicular to the direction of propagation. Thus, light waves are transverse. The oscillations of $\mathbf{E}$ and $\mathbf{B}$ are in phase. At time $t = 0$, the phase at the origin $(x = y = z = 0)$ is φ_0.

It is useful to have a picture of the wave described by Eq. (10.9). This is given in Fig. 10.7 which shows the electric and magnetic fields at one instant of time. The fields $\mathbf{E}$ and $\mathbf{B}$ vary as z is changed. Moving along z by the wavelength λ or an integral multiple of λ brings them back to their original values. Notice that the fields do not depend on x or y in Eq.(10.9), but only on z. The planes parallel to the $x-y$ plane in Fig. 10.7 are therefore wavefronts, *i.e.*, the phase of the wave stays constant as we move along such a plane, varying x and y. As you know from your earlier study of waves (Chapter 13, XI Class), they move forward by a distance λ in the period of oscillation T. The velocity is given by

$$v = \frac{\lambda}{T} = \lambda\nu. \quad (10.10)$$

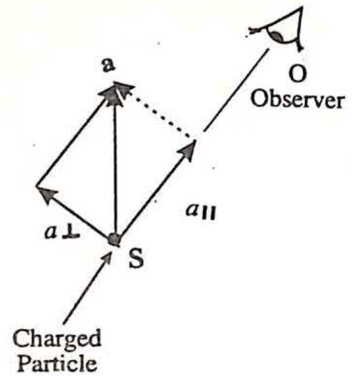


Figure 10.8: A diagram showing the acceleration vector **a** of a radiating charged particle at S. It is resolved into components $\mathbf{a}_{\|}$ and $\mathbf{a}_{\perp}$, parallel and perpendicular respectively to the direction SO pointing to an observer at O. The observer sees radiation with electric field **E** parallel to $\mathbf{a}_{\perp}$, of strength proportional to $|\,\mathbf{a}_{\perp}\,|$.

10.3.2 Charged particles and electromagnetic waves: emission, absorption and scattering

You have already learnt that radio waves are produced by rapidly oscillating electrical currents (Chapter 9). This general principle is true for all electromagnetic waves and thus light as well. Since moving charges make up a current, we can state it as follows. *An accelerated charge radiates electromagnetic waves* (see Section 9.4). For example, the light from the hot filament of a bulb is emitted by the moving electrons inside the metal. These are accelerated by forces from the positive nuclei and other electrons. The same general principle holds for light from the sun. In such processes energy is transferred from the charged particle to the radiated electromagnetic fields. This is known as *emission*.

Fig. 10.8 shows the acceleration vector **a** of a charge at S. The radiation is being observed by an observer at O. Resolve **a** into $\mathbf{a}_{\|}$ along the direction SO of the observer and $\mathbf{a}_{\perp}$ perpendicular to this direction. Because electromagnetic waves are transverse, the component $\mathbf{a}_{\|}$ does not contribute to the radiation received at O. The electric vector of the light emitted toward this observer is parallel to $\mathbf{a}_{\perp}$. The strength is proportional to $|\,\mathbf{a}_{\perp}\,|$, that is, the magnitude of the transverse component of the acceleration. We will use this principle later in discussing polarised light.

Conversely, an incoming electromagnetic wave can produce acceleration in a charged particle. If the magnitude of the velocity of the particle increases, energy is gained by the charge and lost by the wave. This is known as *absorption*. For example, a dark (non reflecting) surface placed in sunlight heats up due to such processes.

Now consider a wave incident on a charged particle at rest. Because the velocity v of the charge is initially zero, the magnetic field B does not produce any force on the charge to start with. The acceleration of the particle is along the electric field.

Once accelerated by an electromagnetic wave a charged particle itself becomes a source. Its radiation is emitted in all directions, not just in the direction of the original wave. This process is called *scattering* of light. For example, the blue light that we receive from the sky is scattered sunlight. Emission, absorption, and scattering are three important processes by which radiation and matter interact.

★ Our discussion of absorption was a simplified one. Actually, the acceleration vector produced by the electromagnetic wave can sometimes be opposite to the velocity vector of the moving charge. In that case the charge loses energy. By conservation, it is clear that the electromagnetic wave gains energy! One needs very special situations to make this happen more often than ordinary absorption. This effect is called *negative absorption* or *stimulated emission* and

underlies the operation of lasers and many electronic amplifiers. ★

In the discussion of scattering, the acceleration was assumed to be approximately parallel to the *electric* vector even if the charge was already moving, so long as its speed v was much less than c, the speed of light. We must check this by calculating the maximum value of the magnetic term in the force on the charge and comparing it to the electric force.

$$\frac{\text{Electric force}}{\text{Magnetic force}} = \frac{qE}{qvB} = \frac{E}{Bv} = \frac{c}{v}.$$

The ratio E/B in a wave is taken from Chapter 9 (Eq. (9.13)). Clearly the electric force is much greater for slowly moving particles ($v << c$). In our further discussion, we will therefore concentrate on the electric field of the wave, though the magnetic field is always present.

10.4 Coherent and incoherent addition of light waves

Our common experience is that when two light bulbs shine simultaneously on the same wall, the intensities add. If light consisted of particles, this would be easy to understand. An area of the wall receives the particles from both the sources, the number of particles is the sum of what each source would give by itself.

For electromagnetic waves, the *electric fields* produced by different sources add. This is called the principle of superposition. The intensity I at a given point is proportional to the *square* of the amplitude of the electric field $\mathbf{E}$ (as mentioned in Section 9.3), *i.e.*, dot product of $\mathbf{E}$ with itself.

$$I = k\mathbf{E}\cdot\mathbf{E}$$

where k is a constant of proportionality.

We will see now that the *intensities* due to two sources *do not add* in general. Let $\mathbf{E}_1$ and $\mathbf{E}_2$ be the fields individually produced at a point by the sources 1 and 2, and I_1 and I_2 their intensities. $\mathbf{E}$ is the electric field and I the intensity when both sources are present. We then have

$$\begin{aligned} I_1 &= k\mathbf{E}_1\cdot\mathbf{E}_1,\ I_2 = k\mathbf{E}_2\cdot\mathbf{E}_2, \\ \mathbf{E} &= \mathbf{E}_1+\mathbf{E}_2, \\ I &= k(\mathbf{E}_1+\mathbf{E}_2)\cdot(\mathbf{E}_1+\mathbf{E}_2) \\ &= k\mathbf{E}_1\cdot\mathbf{E}_1+k\mathbf{E}_2\cdot\mathbf{E}_2+2k\mathbf{E}_1\cdot\mathbf{E}_2 \\ &= I_1+I_2+2k\mathbf{E}_1\cdot\mathbf{E}_2 \qquad (10.11) \end{aligned}$$

This Eq.(10.11) shows that the intensity observed I equals the sum of the individual intensities I_1+I_2, plus one more term

$$I_{12} = 2k\mathbf{E}_1\cdot\mathbf{E}_2. \qquad (10.12)$$

This third term is called the *interference term*. Being a dot product of two vectors, it can be positive, negative, or zero.

The calculation so far was for a given instant of time. But light waves oscillate with a period of the order of 10^{-15} seconds. Intensity measurements are usually made over times much longer than the period. To compare with such experiments, we must calculate the *average* of the interference term $2k\mathbf{E}_1\cdot\mathbf{E}_2$ for a time interval consisting of many periods. To do this calculation, we need two simple results from trigonometry:

(i) The average of $\cos x$ is zero if the angle x varies uniformly over the full range 0 to 2π. Looking at the graph of $\cos x$, we can already guess that the positive and negative parts will cancel each other. Using $\cos(x\pm\pi) = -\cos x$, we prove that for every angle x, there is an angle $x+\pi$ (or $(x-\pi)$ if x is greater than π) for which the cosine is minus that for x. The average will clearly be zero.

(ii) The average of $\cos^2 x$, when x varies uniformly over 0 to 2π, is $1/2$. To see this, we write

$$\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x \qquad (10.13)$$

and use the earlier result that the average of $\cos 2x$ is zero. (We have used these results earlier in Chapter 8 on time varying currents, see for example, Section 8.2).

We now consider the interference term I_{12} in the case when the two electric fields are parallel to each other along a unit vector $\hat{\mathbf{n}}$. Let $\mathbf{E}_1(t)$ and $\mathbf{E}_2(t)$ have amplitudes a_1 and a_2, angular frequencies $\omega_1 = 2\pi\nu_1$ and $\omega_2 = 2\pi\nu_2$ and initial phases, (at time $t = 0$) of ϕ_1 and ϕ_2. We write

$$\begin{aligned}\mathbf{E}_1(t) &= \hat{\mathbf{n}} a_1 \cos(\omega_1 t + \phi_1)\\ \mathbf{E}_2(t) &= \hat{\mathbf{n}} a_2 \cos(\omega_2 t + \phi_2)\end{aligned}$$

$$\begin{aligned}I_{12} &= 2k\mathbf{E}_1 \cdot \mathbf{E}_2\\ &= 2k\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} a_1 a_2 \cos(\omega_1 t + \phi_1)\\ &\quad \times \cos(\omega_2 t + \phi_2)\\ &= k a_1 a_2 \{\cos[(\omega_1 - \omega_2)t + \phi_1 - \phi_2]\\ &+ \cos[(\omega_1 + \omega_2)t + \phi_1 + \phi_2]\}\end{aligned} \qquad (10.14)$$

Averaging I_{12} over many cycles, we find the second term inside the curly brackets becomes zero. The average of the first term is also zero if $\omega_1 \neq \omega_2$. We conclude that:

(a) The *intensities* of two beams with *different* frequencies add when we average over time.

Suppose the two frequencies are the same, we then have

Average of I_{12}

$$= k a_1 a_2 \cos(\phi_1 - \phi_2). \qquad (10.15)$$

The interference term is proportional to the cosine of the phase difference. However, when we speak of two different sources of light of the same frequency, say two sodium lamps, the motions of the charges in them are independent. The phase difference ($\phi_1 - \phi_2$) will not have a fixed or stable value during the measurement, and the average of I_{12} will be zero.

(b) The intensities from two *independent* sources, even of the same frequency, add.

(c) To see interference, we need two sources with the same frequency *and* with a stable phase difference. Such a pair of sources is called *coherent*. As we will see in the next section, such a pair of sources is obtained in practice by deriving both from a single source.

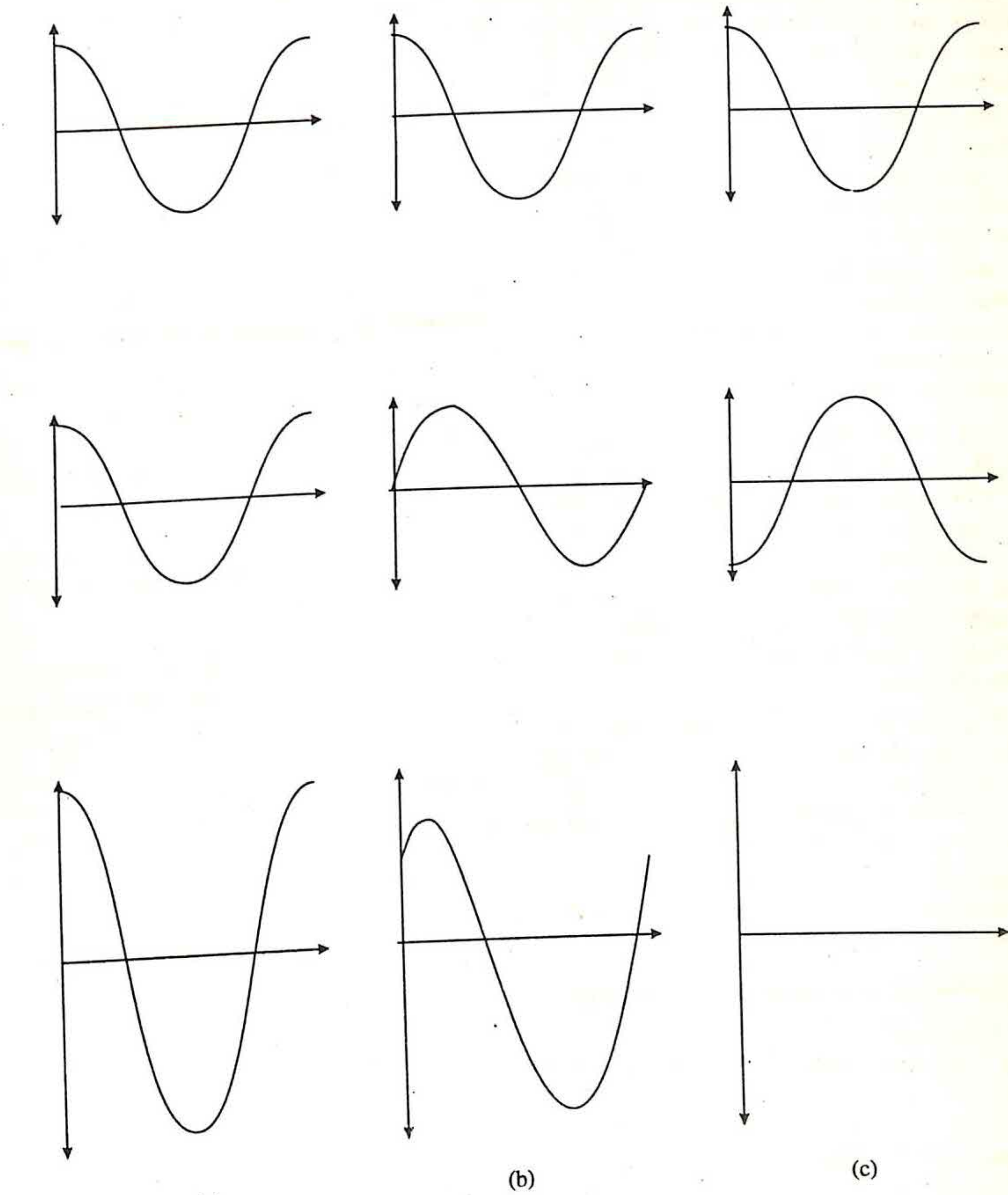
Now we can understand why $I = I_1 + I_2$ in most experiments - the two sources have a phase difference which is not stable but itself varies over 0 to 2π. The interference term averages to zero. We call two such sources *incoherent*.

★ With sound waves, or at radio frequencies, one has much more control over the phases even of two independent sources and can make intensity measurements over short periods. For example, we get beats between two tuning forks even when the frequencies are slightly different. There is no *fundamental* difference between electromagnetic waves with frequencies in the radio band and visible light. With modern techniques such as the use of lasers, it is possible to observe interference and beats between two independent sources even for visible light. ★

10.5 Interference

10.5.1 Superposition of two waves

The ideas of superposition and interference can be simply visualised using water waves. Imagine two sticks S_1 and S_2 being moved periodically up and down in an identical



(a) (b) (c)

Figure 10.9: Graphs of water level versus time. First row: as produced by the first source a
Second row: as produced by the second source alone. Third row: combined effect of both source
in phase, (b) phase difference of 90° and (c) phase difference of 180°.

fashion in a trough of water. They produce two water waves whose phase difference is stable, *i.e.*, they are coherent. The oscillations in water level produced by the two sources will be in phase at any point P whose distances S_1P and S_2P from the two sources are equal. Graphs of displacement versus time as produced by each source separately are shown in Fig. 10.9a. The lowermost graph shows the resultant displacement which is twice that of one source. This means that the intensity is four times that produced by one source. This is called *constructive interference.*

Fig. 10.9c shows a case when the path from one of the sources to a given point is half a wavelength longer than the other. This produces a phase difference of 180° or π radians between the two sources. Compare the first two graphs. The total displacement is zero in this case (when the amplitudes are equal). This situation is called *destructive interference.*

Fig. 10.9b shows an intermediate case when the two oscillations have a phase difference of 90°. In this case the resultant intensity of the combined oscillation is twice that due to one source and its phase is 45° which is midway between the phases of two sources.

Example [illegible] Combine [illegible]tions of equal amplitude and 90° phase difference.

Answer: If we take the first to be $\cos\omega t$, the second is $\cos(\omega t - \frac{\pi}{2}) = \sin\omega t$. The superposition is

$$\begin{aligned}
&\sin\omega t + \cos\omega t \\
&= \sin(\omega t) + \sin\left(\frac{\pi}{2} - \omega t\right) \\
&= 2\sin\left[\frac{1}{2}[\omega t + \frac{\pi}{2} - \omega t]\right] \times \\
&\qquad \cos\left[\frac{1}{2}[\omega t + \omega t - \frac{\pi}{2}]\right] \\
&= 2\sin\frac{\pi}{4}\cos\left(\omega t - \frac{\pi}{4}\right) \\
&= \sqrt{2}\cos\left(\omega t - \frac{\pi}{4}\right)
\end{aligned}$$

The intensity is proportional to $(\sqrt{2})^2 = 2$ and the phase lags the 'cos'vibration by 45°.

For the purpose of calculating phase differences a path of a whole number of wavelengths makes no difference. We can, therefore, state the following general results for two sources in phase.

(a) A path difference equal to an integer n times the wavelength gives constructive interference.

(b) A path difference of half a wavelength plus an integer times the wavelength gives destructive interference.

10.5.2 Young's experiment

An experiment first carried out in the year 1802 by the English scientist Thomas Young is one of the most beautiful demonstrations of the wave nature of light. Two slits S_1 and S_2 are made in an opaque screen, parallel to each other and very close (Fig. 10.10). These two are illuminated by another narrow slit S which is in turn lit by a bright source (all early experiments used the sun). As we will discuss in more detail in the next section on diffraction, light waves spread out from S and fall on both S_1 and S_2. S_1 and S_2 then behave like two coherent sources - the two sticks in our water wave example. Thus, the two coherent light waves are derived from the same original source. In this way, any phase change in S occurs in both S_1 and S_2.

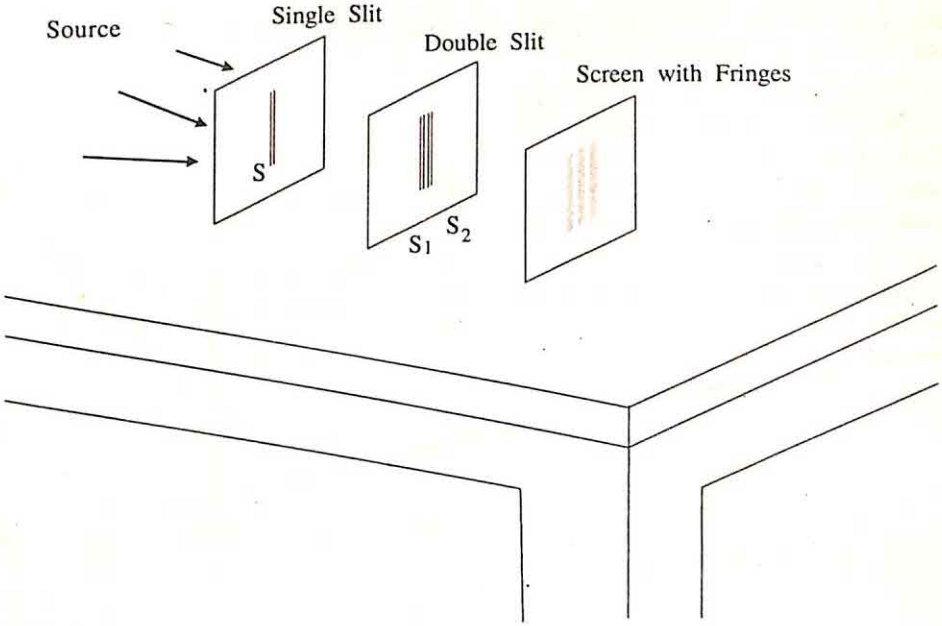


Figure 10.10: A schematic view of Young's double slit interference experiment. The widths of the slits, their separation and the fringes are not to scale. They have been shown bigger for clarity.

The phase difference $\phi_1 - \phi_2$ between S_1 and S_2 is unaffected and remain stable.

Light now spreads out from both S_1 and S_2 and falls on a screen. It is essential that the waves from the two sources overlap on the same part of the screen. If one slit is covered up, the other produces a wide smoothly illuminated patch on the screen (which we will study in detail in the section on diffraction). But when both slits are open, the patch is seen to be crossed by dark and bright bands called *interference fringes*. A part of the screen which received light from either S_1 or S_2 alone can become dark when both slits are open! This is a clear case of destructive interference and hence a proof of the wave nature of light.

Fig. 10.11 shows graphs of how the intensity on the screen varies as one moves along a line perpendicular to the slits. The thick line shows the case when both slits are open and the fringes are formed. For comparison, the intensity received at the screen from a single slit is also shown as a dotted line. These graphs show that the intensity minimum is zero and the maximum is four times the contribution of a single slit. This is easily understood using the principle of superposition. For simplicity, we take the amplitude of the electric field produced by each slit to be E. We get zero electric field when the interference is destructive. For constructive interference, the field is $2E$ and the intensity proportional to $4E^2$. Note that the *average* intensity of the maximum and minimum is $2E^2$, just the sum of what the two slits produce. The phenomenon of interference therefore redistributes the energy keeping the total fixed.

Path differences

To understand the fringe pattern in more detail, we need to calculate the path lengths between the slits S_1 and S_2, and points on

the screen. Fig. 10.12a shows three parallel planes. The first contains the slit S, the second the two slits S_1 and S_2 and the third a typical point P where the light falls on the screen. The point O is midway between S_1 and S_2. The line perpendicular to all the planes, and passing through O, meets the source plane at O_2 and the screen at O_1. These two points are used as origins to measure the positions of S and P.

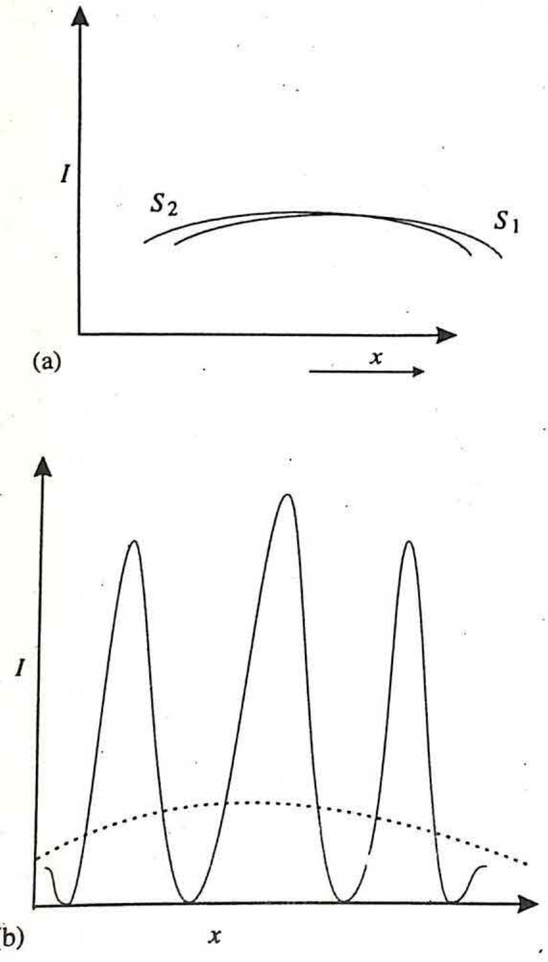
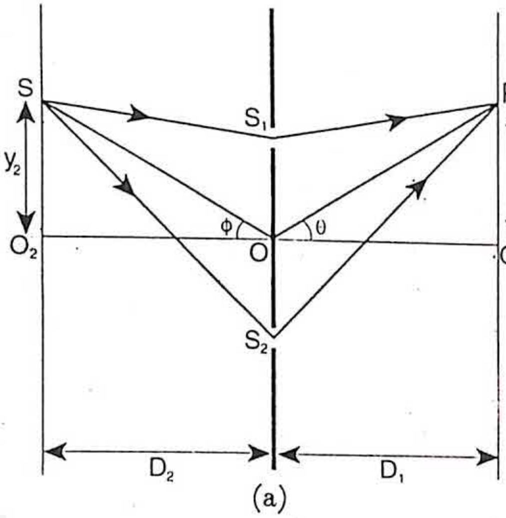


Figure 10.11: Intensity I of light plotted against position x on the screen in Young's experiment: (a) Slit S_1 or slit S_2 individually opened, (b) Both slits open showing interference fringes. The dashed curve is the single slit intensity for comparison.

$$O_1P = y_1\,;\;\; O_2S = y_2 \qquad (10.16)$$



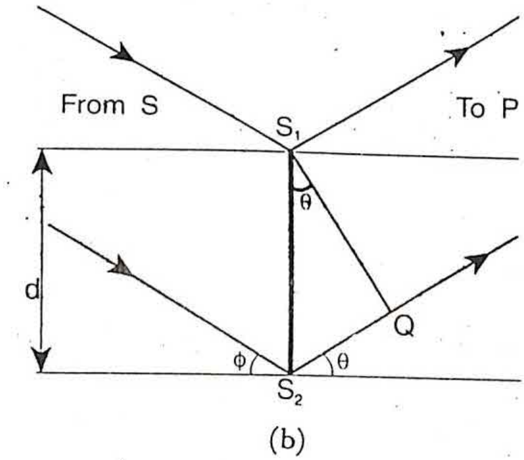


Figure 10.12: Geometry of path differences in Young's experiment. (a) Coordinate system defining the positions of the slit S, double slit S_1S_2 and point of observation on the screen, P. (b) Magnified view of the paths reaching the two slits from the source S and travelling to the screen at P.

To start with, assume that S is equidistant from S_1 and S_2 (that is, S coincides with O_2 so that the angle ϕ in Fig.12b is zero). We need to calculate the path difference, $S_2P -$

S_1P. As shown in Fig. 10.12b, the two lines S_1P and S_2P are nearly parallel since the distance $S_1S_2 = d$ is much less than $OO_1 = D_1$. The angle that these two lines make with the normal to the screen is denoted by θ. A perpendicular S_1Q is dropped from S_1 to the line S_2P. The angle subtended at P by S_1Q is very small. We can therefore think of S_1Q as the arc of a circle centered at P with radius PS_1. The length PQ then is very nearly equal to PS_1. The desired path difference equals

$$\begin{aligned} S_2P - S_1P &= S_2P - QP \\ &= S_2Q = d\sin\theta \\ &\approx d\tan\theta \approx \frac{dy_1}{D_1}. \end{aligned} \qquad (10.17)$$

In deriving Eq.(10.17) we have used the result for small angles that $\sin\theta \approx \theta \approx \tan\theta$ (for small θ) $\approx y_1/D$.

The condition for constructive interference reads

$$\begin{aligned} S_2P - S_1P &= \frac{dy_1}{D_1} = n\lambda \\ \text{i.e., } y_1 &= n\frac{D_1\lambda}{d}. \end{aligned} \qquad (10.18)$$

Putting $n = 0$, we have a central bright fringe corresponding to zero path difference at $y_1 = 0$. The separation between two successive maxima is found by subtracting the values of y_1 corresponding to successive values of n. We denote this by Δy_1

$$\Delta y_1 = \frac{D_1\lambda}{d}(n+1-n) = \frac{D_1\lambda}{d}. \qquad (10.19)$$

This separation Δy_1 between two adjacent bright (or dark) fringes subtends an angle $\Delta\theta$ at the centre O of the double slit.

$$\Delta\theta = \frac{\Delta y_1}{D_1} = \frac{\lambda}{d}. \qquad (10.20)$$

The angular separation of the fringes is just λ/d independent of the position of the screen. This is the increase in θ needed in Eq.(10.17) to increase the path difference by λ.

Young's experiment therefore gives a direct way of measuring λ, the wavelength of light. If the original source is an electric lamp or sunlight, a filter or prism is needed to separate the colour of interest.

Example 10.2: Two slits are made one millimeter apart and the screen is placed one metre away. What is the fringe separation when blue - green light of wavelength 500nm is used?

Answer:

$$\text{Fringe spacing} = \frac{D\lambda}{d} = \frac{1\times 5\times 10^{-7}}{1\times 10^{-3}}$$

$$\mathrm{m} = 5\times 10^{-4}\,\mathrm{m} = 0.5\,\mathrm{mm}$$

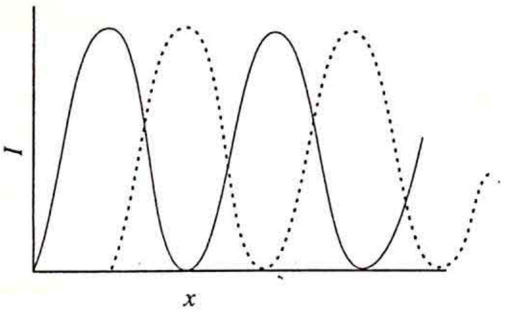
For simplicity we have so far placed the first slit S at $y_2 = 0$. This produced zero path difference at $y_1 = 0$. When we have a source at y_2, the lines SS_1 and SS_2 make an angle $\varphi = y_2/D_2$ with the x-axis (Fig. 10.12b). At the central fringe, the total path difference is zero, that is

$$\begin{aligned} &SS_2 + S_2P - SS_1 - S_1P \\ &= (S_2P - S_1P) + (S_2S - S_1S) = 0; \end{aligned}$$

$$\frac{dy_2}{D_2} + \frac{dy_1}{D_1} = 0$$

$$\text{that is } \frac{y_2}{D_2} = -\frac{y_1}{D_1} \text{ or } \theta = -\varphi. \qquad (10.21)$$

The meaning of Eq.(10.21) is that the angular movement θ of the central fringe is equal and opposite to the angular movement φ of the source. Both the angles are measured at the centre of the double slit. More simply, the first slit, the centre of the double slit and the central fringe lie in a straight line as the source is moved.



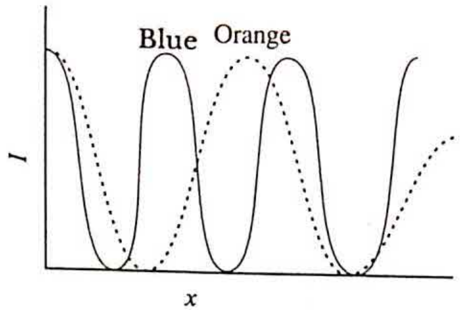


Figure 10.13: Two factors affecting the visibility of interference fringes. (a) Each incoherent part of the source produces its own fringe pattern. Because their maxima and minima are different, the fringe contrast is poor. (b) Different wavelengths present in the source add incoherently, each producing its own fringe pattern. The longer wavelengths like orange produce the dashed fringe pattern which has a larger spacing than that for a shorter wavelength like blue (solid line). The different patterns agree at the central fringe (zero path difference). At larger path differences the fringes become coloured and then fainter.

★ *Visibility of the interference fringes*: Suppose one uses a source (like the filament of a bulb) in place of the narrow slit S. Let this have a size Δy_2. Each part of the source produces a fringe pattern on the screen. Since these different parts are incoherent (independent sources), we can add the intensities of the different fringe patterns on the screen. Their central fringes will be spread out over an angle $\Delta\theta = \Delta y_2/D_2$. The effect of adding just two equal fringe patterns is shown in Fig. 10.13a. If the maximum of one falls on the minimum of the other then there is no intensity variation in the total. This means the fringes are no longer visible.

If the angle $\Delta\theta$ is much less than λ/d (the angular separation of two maxima) then the fringes from the different parts of the source have maxima and minima at nearly the same place on the screen. We then see the full fringe pattern in the total intensity. But if $\Delta\theta$ is much greater than λ/d, then the maxima of the different patterns which we are adding are spread out by much more than the fringe spacing. The fringe pattern will be washed out. In example (10.2), the spacing between bright fringes was 0.5 mm subtending an angle of 1/2000 radian at the slits. Any source subtending a much larger angle then 5×10^{-4} radian, i.e., $100''$, at the slits will not give clear fringes.

We also have to consider the presence of many wavelengths in the incident light covering a range $\Delta\lambda$ around λ. When we speak of a monochromatic source, it just means $\Delta\lambda$ is small compared to λ, not that $\Delta\lambda$ is zero! The central fringe $n = 0$ occurs at $\theta = 0$ for all wavelengths. But take the example of the $n = 2$ fringe for blue light of $\lambda = 450$ nm. The path difference is 900 nm. For orange light of $\lambda = 600$ nm, this path difference is one and a half wavelengths and there is destructive interference. Fig. 10.13b shows two fringe patterns for different wavelengths. We add the intensities since different wavelengths are incoherent. The central bright fringe is the same for all wavelengths. As the path difference increases the fringes get washed out because the maximum for one wavelength happens to be a minimum for some other wavelength. By making the range of wavelengths $\Delta\lambda$ narrower, one can get a larger number of fringes. For example, let the range be 450 nm to 500 nm. It is only for $n = 5$ for $\lambda = 450$ nm that we have $n = 4\frac{1}{2}$ for $\lambda = 500$ nm. About 5 fringes are seen distinctly on either

side of the central maximum. ★

10.5.3 The colours of thin films

In the rainy season, one sometimes sees oil from some motor vechicle spilt on the road, spreading out to form a thin layer on water. Such a thin layer often shows brilliant colours even when illuminated with white light. Another thin film showing brilliant colours is a soap bubble. In this case the film slowly becomes thinner near the top of the bubble as the water drains down. It is then found that this very thin layer near the top looks dark in reflected light. These colours shown by thin films can be understood as interference between two beams. One is reflected from the top surface of the film and the other from the bottom. As an example, a path difference of 1000 nm leads to destructive interference in the violet side of the spectrum $\lambda = 400$ nm, but constructive interference in the blue green region ($\lambda = 500$ nm). The reflected light will therefore be coloured. The connection between film thickness and colour was studied experimentally by Newton. However, he could not explain it satisfactorily as he did not fully believe in wave ideas.

One might think that a very thin film would give zero path difference and hence constructive interference at all wavelengths. But what is observed, for example, in a thin soap film, is actually destructive interference. The reason is that there is a phase difference of π between a ray in air reflected from water, and a ray travelling in water reflected from air. This result explains why there is destructive interference for a very thin film. This is also what we should expect from a different argument. For a very thin film, the number of molecules available to scatter the light back is very small, so such a film must look dark in reflected light. This extra phase change of π is used in the problems section, where many examples of thin film interference are given.

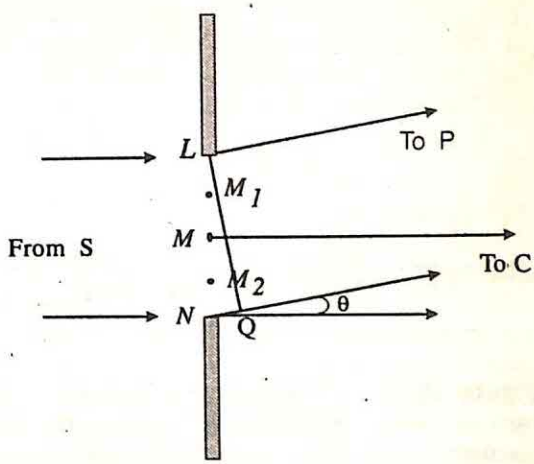


Figure 10.14: The geometry of path difference for diffraction by a single slit.

10.6 Diffraction

10.6.1 The single slit

In the discussion of Young's experiment, we stated that a single narrow slit acts as a new source from which light spreads out. Early experimenters had noticed that light spreads out from narrow holes and slits. It seems to turn round corners and enter regions where we would expect a shadow. These effects, known as *diffraction*, can only be properly understood using wave ideas. After all, you are hardly surprised to hear sound waves from some one talking around a corner!

When the double slit in Young's experiment is replaced by a single narrow slit, a broad pattern with a central bright region is seen. On both sides, there are alternate dark and bright regions, the intensity becoming weaker away from the centre (graph in Fig. 10.15). To explain this, go to Fig. 10.14 which shows a parallel beam of light

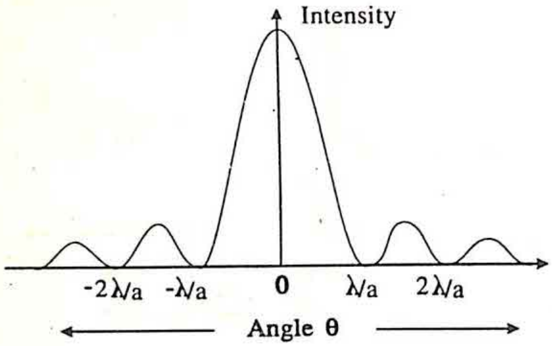


Figure 10.15: The variation of intensity with angle in single slit diffraction. The first secondary maximum is only 4% of the central maximum, so it is not to scale in the figure.

falling normally on a single slit LN of width a. The diffracted light goes on to meet a screen. The midpoint of the slit is M. A straight line through M perpendicular to the slit plane meets the screen at C. We want the intensity at any point P on the screen. As before, straight lines joining P to the different points L, M, N, etc., can be treated as parallel, making an angle θ with the normal MC.

The basic idea is to divide the slit into much smaller parts, and add their contributions at P with the proper phase differences. We are treating different parts of the wavefront at the slit as secondary sources. Because the incoming wavefront is parallel to the plane of the slit, these sources are in phase.

The path difference LP-NP between the two edges of the slit can be calculated exactly as for Young's experiment. From Fig. 10.14, $\mathrm{LP} - \mathrm{NP} = \mathrm{NQ} = a\sin\theta \approx a\theta$. Similarly, if two points $\mathrm{M_1}$ and $\mathrm{M_2}$ in the slit plane are separated by y, the path difference $\mathrm{M_1P} - \mathrm{M_2P} \approx y\theta$. We now have to sum up equal, coherent contributions from a large number of sources, each with a different phase. This calculation was made by Fresnel but uses integral calculus, so we omit it here. The main features of the diffraction pattern can be understood by simple arguments, given below.

At the central point C on the screen, the angle θ is zero. All path differences are zero and hence all the parts of the slit contribute in phase. This gives maximum intensity at C. Now choose an angle θ such that the path difference between the edges L and N at P is λ. This angle is given by

$$a\theta = \lambda,\ \theta = \frac{\lambda}{a}. \tag{10.22}$$

Further, divide the slit into two equal halves LM and MN each of size $a/2$. For every point $\mathrm{M_1}$ in LM, there is a point $\mathrm{M_2}$ in MN such that $\mathrm{M_1M_2} = a/2$. The path difference between $\mathrm{M_1}$ and $\mathrm{M_2}$ at P$= \mathrm{M_1P} - \mathrm{M_2P} = \theta \cdot a/2 = \lambda/2$ for the angle chosen. This means that the contributions from $\mathrm{M_1}$ and $\mathrm{M_2}$ are 180° out of phase and cancel in the direction $\theta = \lambda/a$. Contributions from the two halves of the slit LM and MN therefore cancel each other. Eq.(10.22) gives the angle at which the intensity falls to zero. One can similarly show that the intensity is zero for $\theta = n\lambda/a$, with n any integer (except zero!). Notice that the angular size of the central maximum increases when the slit width a decreases.

Let us now consider an angle $\theta = 3\lambda/2a$ which is midway between two of the dark fringes $\theta_1 = \lambda/a$ and $\theta_2 = 2\lambda/a$. If we take the first two thirds of the slit, the path difference between the two ends would be

$$\frac{2}{3}a \times \theta = \frac{2a}{3} \times \frac{3\lambda}{2a} = \lambda. \tag{10.23}$$

The first two thirds of the slit can therefore be divided into two halves which have a $\lambda/2$ path difference. The contributions of these two halves cancel in the same manner as described earlier. Only the remaining one third

of the slit contributes to the intensity at a point between the two minima. Clearly, this will be much weaker than the central maximum (where the entire slit contributes in phase). One can similarly show that there are maxima at $(n + 1/2)\lambda/a$ with $n = 2, 3$ etc. These become weaker with increasing n, since only one fifth, one seventh; .... etc., of the slit contributes in these cases.

The graph in Fig. 10.15 shows how the intensity on the screen varies with the angle θ. It shows the central maximum at $\theta = 0$, zero intensity at $\theta = \pm n\lambda/a (n \neq 0)$ and secondary maxima at $\theta = \pm(n + 1/2)\lambda/a$ $(n \neq 0)$.

We now go back to the intensity produced by each slit in Young's experiment when the other is blocked. We have a central maximum which is broad if the slit is made narrow. The graphs given in Fig. 10.11 show only a small portion of the central maximum of the single slit diffraction pattern.

Example 10.3: In example 10.2, what should the width of each slit be to obtain 10 maxima of the double slit pattern within the central maximum of the single slit pattern?

Answer:

$$\text{We want} \quad 10\frac{\lambda}{d} = 2\frac{\lambda}{a}$$

$$\text{that is} \quad a = \frac{d}{5} = 0.2\,\text{mm}.$$

Notice that the wavelength and distance of the screen do not enter.

Our discussion of Young's experiment and of single slit diffraction has assumed that the screen on which the fringes form is at a large distance. The two or more paths from the slits to the screen were treated as parallel. This situation also occurs when we place a converging lens after the slits and place the screen at the focus. Parallel paths from the slit are combined at a single point on the screen. This arrangement is often used since it gives more intensity than placing the screen far away.

★ *Seeing the single slit diffraction pattern*: It is surprisingly easy to see the single slit difraction pattern for oneself. The equipment needed can be found in most homes - two razor blades and one clear glass electric bulb preferably with a straight filament. One has to hold the two blades so that the edges are parallel and have a narrow slit in between. This is easily done with the thumb and forefingers (Fig. 10.16). Keep the slit right in front of the eye and close to it, parallel to the filament. Use spectacles if you normally do. With slight adjustment of the width of the slit and the parallelism of the edges, the pattern should been seen with its bright and dark bands. Since the position of all the bands (except the central one) depends on wavelength, they will show some colours. Using a filter for red or blue will make the fringes clearer. With both filters available, the wider fringes for red compared to blue can be seen.

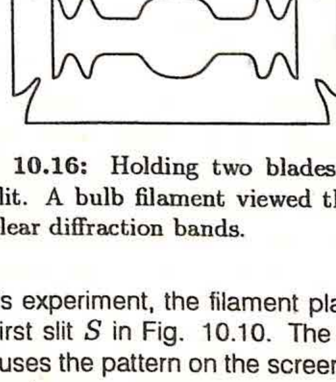


Figure 10.16: Holding two blades to form a single slit. A bulb filament viewed through this shows clear diffraction bands.

In this experiment, the filament plays the role of the first slit S in Fig. 10.10. The lens of the eye focuses the pattern on the screen (the retina of the eye).

With some effort, one can cut a double slit in an aluminium foil with a blade. The bulb filament can be viewed as before to repeat Young's experiment. In daytime, there is another suitable bright source subtending a small angle at the eye. This is the reflection of the sun in any shiny convex surface (e.g., a cycle bell). Do not try direct sunlight - it can damage the eye and will not give fringes anyway as the sun subtends $(1/2)°$

★

10.6.2 The diffraction grating

We have already seen more than once that the conditions for constructive and destructive interference depend on the wavelength and give rise to colours in the fringes except for zero path difference. This principle is used in a device called a diffraction grating which separates light into different colours. It can therefore be used in much the same way as a prism. In its simplest form, it consists of a large number N of single slits separated by d, the total width w being Nd. Fig. 10.17a shows how one gets constructive interference between waves from all the slits in a direction θ to the normal. Since the slits are equally spaced, it is enough to ensure that two successive slits are in phase with a path difference of $n\lambda$. The condition for this reads

$$d\sin\theta = n\lambda;\ \theta \approx \frac{n\lambda}{d} \qquad (10.24)$$

Clearly, different colours are sent in different directions. For a fixed wavelength λ, there are diffracted beams corresponding to $n = 1, 2, 3$ etc. These are called the first order, second order, etc., spectra. The graph in Fig. 10.17b shows how the intensity varies with angle θ. As with a single slit, it is convenient to use a lens and focus the pattern on the photographic plate, electronic detector or screen.

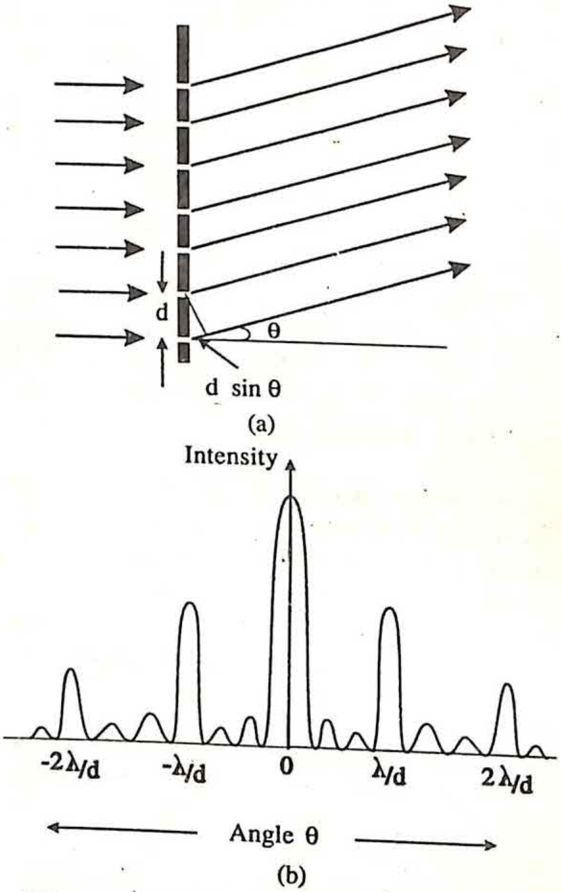


Figure 10.17: (a) The geometry of path differences for a diffraction grating. Every successive pair has the same path difference of $d\sin\theta$. (b) Intensity versus angle for monochromatic light incident on a grating. Note the zero, first, second order maxima. For an actual grating the angular width of each maximum is very much less than the spacing between maxima. (Example 10.6).

Example 10.4: A diffraction grating one cm wide has 1000 lines and is used in third order. What are the difraction angles for violet and orange light? What is the angular size of the diffraction maximum for monochromatic light? The wavelengths for

violet and orange are 400 nm and 600 nm.

Answer: For third order $n = 3$, and we have $d = 10^{-5}$ m. We denote the diffraction angles for violet and orange by θ_v and θ_0

$$\begin{aligned}\theta_v &= \frac{3 \times 4 \times 10^{-7}\ \text{m}}{10^{-5}\ \text{m}}\ \text{rad} \\ &= 12 \times 10^{-2}\ \text{rad} \approx 6.9°, \\ \theta_0 &= 18 \times 10^{-2}\ \text{rad} \approx 10.3°.\end{aligned}$$

The spectrum is thus spread over an angle of about 3.4°.

At a maximum, we have $\theta = 3\lambda/d$. The path difference between the first and last slit in the grating is an integral number of wavelengths. Let us increase θ so that an extra path difference of λ is introduced between the two ends of the grating, separated by w. The change in θ required to do this is denoted by $\Delta\theta$.

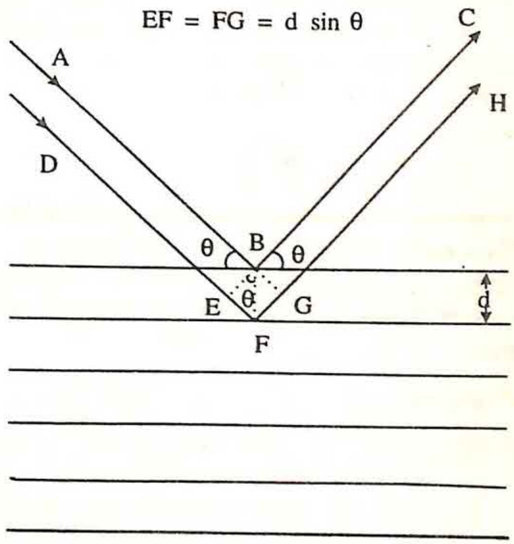
$$w\Delta\theta = \lambda, \ \ \Delta\theta = \frac{\lambda}{w}.$$

Because of the 360° extra phase across the grating, we can again divide it into two halves so that there is a 180° phase difference between slits separated by $w/2$. (We used this argument earlier to find the first zero of the single slit pattern). Therefore we get zero intensity at an angle $\Delta\theta$ away from the maximum.

$$\begin{aligned}\Delta\theta &= \frac{\lambda}{w} \\ &= \frac{4 \times 10^{-7}\ \text{m}}{1 \times 10^{-2}\ \text{m}} = 4 \times 10^{-5}\ \text{radians} \\ &= 2.3 \times 10^{-3}\ \text{degrees} \\ &\quad \text{for violet light}\end{aligned}$$

Notice the maximum is very narrow. A diffraction grating can therefore resolve (*i.e.,* show as separate) two wavelengths which are very close to each other.

Example 10.5: A beam with wavelength λ falls on a stack of partially reflecting planes with separation d. What angle θ should it make with the planes so that the beams reflected from successive planes interfere constructively?



Answer: The extra path traversed by the beam DEFGH compared to the beam ABC equals EF + FG = 2d sin θ. One gets constructive interference when this is an integer times the wavelength $n\lambda = 2d\sin\theta$. This result (known as Bragg's law) is important for the study of crystals by X-rays (Chapter 12).

One common object which can act as a diffraction grating is a long playing record (or compact disc) which has many closely spaced grooves. One only has to view the reflection of a tube light (parallel to the grooves) in such a record to see the separation of colours.

10.6.3 The validity of ray optics

An aperture (*i.e.*, slit or hole) of size a illuminated by a parallel beam sends diffracted light into an angle of approximately $\approx \lambda/a$. This is the angular size of the bright central maximum. In travelling a distance z, the diffracted beam therefore acquires a width $z\lambda/a$ just due to diffraction. It is interesting to ask at what value of z the spreading due to diffraction become greater than the size a of the aperture. We want

$$\frac{z\lambda}{a} > a, \text{ i.e., } z > \frac{a^2}{\lambda} \tag{10.25}$$

Example 10.6: For what distance is ray optics a good approximation when the aperture is 3 mm wide and the wavelength 500 nm

Answer:

$$\frac{a^2}{\lambda} = \frac{(3 \times 10^{-3}\ \text{m})^2}{5 \times 10^{-7}} = 18\text{m}$$

This example shows that even with a small aperture, diffraction spreading can be neglected for rays many meters in length. Thus ray optics is valid in many common situations.

★ The distance a^2/λ in Eq.(10.25) is called the Fresnel distance z_F.

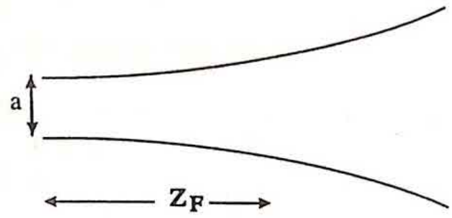
The same formula can be viewed in a different way. If we want a beam to travel a distance z without too much broadening by diffraction, we must have

$$z_F > z, \text{ i.e., } \frac{a^2}{\lambda} > z,$$

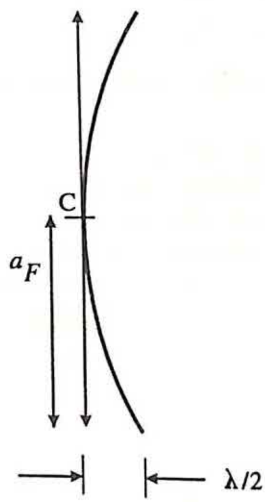
$$a > \sqrt{\lambda z} \equiv a_F \tag{10.26}$$

For a given value of z, the radius a should be greater than $\sqrt{\lambda z}$. This quantity is called the Fresnel zone radius and denoted by a_F

$$a_F = \sqrt{\lambda z}. \tag{10.27}$$



(a)



(b)

Figure 10.18: (a) A parallel beam of size a starts to diverge appreciably after travelling a distance z_F. (b) In calculating the field at a distance z, the phase varies by $\lambda/2$ over a region of size a_F. Outside this the phase varies rapidly and cancellation occurs.

The physical meaning of z_F and a_F is illustrated in Figs (10.18a) and (10.18b). In Fig. 10.18a we have a paralled beam (plane wave) limited to size a by an aperture. For small distances z, the beam diameter continues to be nearly a. When the beam travels a distance z_F from the aperture, it broadens significantly due to diffraction. When z is much greater than z_F, we have a spherical wave occupying an angle λ/a.

In Fig. 10.18b, we have a point P at a distance z from a plane on which C is the nearest point to P, i.e., CP is perpendicular to the plane. A circle of radius a_F is drawn on the plane with C as centre. A point D on this circle has a

distance from C which is greater than CP by half a wavelength.

$$\begin{aligned} a_F^2 = CD^2 &= DP^2 - CP^2 \\ &= \left(z + \frac{\lambda}{2}\right)^2 - z^2 \\ &\approx z\lambda \left(\text{neglecting } \frac{\lambda^2}{4}\right) \end{aligned}$$

Clearly, points well inside this circle of radius a_F all contribute almost in phase with C. Points will outside this circle contribute with a phase which is large and varies rapidly. Fresnel introduced the concept of a *half period zone* of radius a_F. We can say, approximately, that the important contribution to the field of the light wave at P comes from this zone within which there is good phase agreement. Contributions from outside tend to cancel each other out. ★

Example 10.7: Two towers are built on hills 50 km apart and the line joining them passes 30 m above a hill halfway in between. What is the longest wavelength of radio waves which can be sent between the towers without serious diffraction effects caused by the central hill?

Answer: We want the Fresnel zone size a_F at the middle hill to be much less than 30 metres. *i.e.*, $\sqrt{25,000\lambda} \ll 30$ which means $\lambda \ll 30^2 \text{ m}^2/25,000 \text{ m} = 0.036$ m. Wavelengths of order 3.6 cm or longer will undergo serious diffraction effects.

10.7 The Doppler effect for light

You have already encountered the Doppler effect for sound waves (Chapter 13, Class XI text book). If the source of sound waves travels toward the observer, the frequency measured becomes higher than if the source were at rest. When the source travels away, the frequency decreases. The Doppler effect is a rather basic property of waves and is naturally expected to occur for light as well. In Fig. 10.19, a source S is shown emitting waves of wavelength λ and frequency $\nu = c/\lambda$, c being the speed of light. The periodic time of the oscillations T equals $1/\nu$. The figure shows successive wavefronts separated by λ. In time t, the source emits t/T cycles. An observer R at rest would also count the same number of cycles t/T. Now consider a moving observer M. Let v_r be the radial component of the velocity of M, that is, component towards the source. This is also the component perpendicular to the wavefronts. In a time t, M would have come closer to S by $v_r t$ and hence crossed $v_r t/\lambda$ additional cycles of the wave which have not reached R. Notice that the component of velocity parallel to the wavefronts does not affect this count. The total number of cycles crossed by M will be

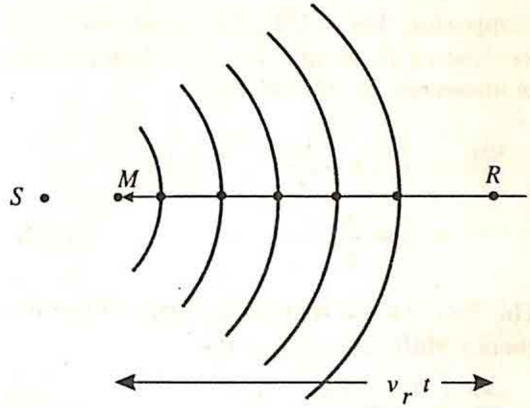


Figure 10.19: The Doppler effect for a stationary source. In a time t, the moving observer M intersects more cycles of the wave than the stationary observer R.

$$\begin{aligned} \frac{t}{T} + \frac{v_r t}{\lambda} &= \frac{t}{T} + \frac{v_r}{c} \times t\frac{c}{\lambda} \\ &= \frac{t}{T}\left(1 + \frac{v_r}{c}\right) \end{aligned} \quad (10.28)$$

Comparing this to t/T, the number of cycles received by R, we get the ratio of frequencies as measured by M and by R

$$\frac{\nu_M}{\nu_R} = \frac{t}{T}\left(1+\frac{v_r}{c}\right) \Big/ \left(\frac{t}{T}\right) = 1+\frac{v_r}{c}. \tag{10.29}$$

This can also be written in terms of the frequency shift $\Delta\nu = \nu_M - \nu_R$

$$\frac{\Delta\nu}{\nu_R} = \frac{\nu_M - \nu_R}{\nu_R} = \frac{v_r}{c}. \tag{10.30}$$

★ In the case of sound waves, one has a different formula when the source is moving. In the time T between the emission of two successive wavefronts, the source moves forward by $v_r T$. The wavelength λ_R emitted by the source at rest is modified to $\lambda_M = \lambda_R - v_r T$. The ratio of frequencies is

$$\begin{aligned}\frac{\nu_M}{\nu_R} &= \frac{\lambda_R}{\lambda_M} = \frac{\lambda_R}{\lambda_R - v_r T} \\ &= \frac{1}{1 - v_r T(\nu_R/c)} \\ &= \frac{1}{1-(v_r/c)} \\ &\quad \text{because } \nu_R T = 1.\end{aligned} \tag{10.31}$$

Notice that this is not exactly the same as Eq.(10.29). However, for small values of (v_r/c), (a source moving much slower than light) powers of v_r/c higher than the first can be neglected. In this case of $v_r/c \ll 1$, we have $1/(1 - v_r/c) \approx 1 + v_r/c$. We conclude that Eqs.(10.29) and (10.31) for the Doppler effect agree for small values of v_r/c.

In the case of light waves travelling in vacuum, a statement that the source or the observer is moving depends on the frame of reference. We expect the ratio of the frequency received to that transmitted to be the same in both cases. Our two Eqs.(10.31) and (10.29) do not agree. The reason is that neither of them is exact. A correct treatment of the Doppler effect for light needs the special theory of relativity, which is not discussed in detail in this book. For completeness, we give the correct relativistic formula without a derivation

$$\frac{\nu_M}{\nu_R} = \frac{1+v_r/c}{\sqrt{1-v^2/c^2}}. \tag{10.32}$$

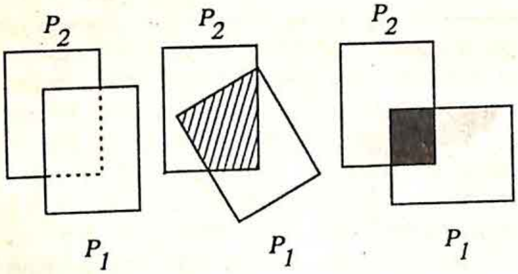
The formulae we have given earlier are correct only for small v/c. ★

A wavelength in the middle of the visible part of the spectrum is shifted to higher frequencies, that is towards blue when source and observer approach each other and to the red when they recede from each other. Astronomers use the terms blue shift and red shift for frequency increase and decrease, (*even when the wavelength is not in the visible spectrum!*) As you will learn in Chapter 15, measurement of Doppler shifts is a powerful tool used by astronomers to study the motions of stars and galaxies. At the earthly level, the Doppler effect of reflected short radio waves is used by police to detect overspeeding vehicles!

10.8 The polarisation of light

10.8.1 Polarisers and analysers

The fact that light consists of transverse waves was discovered in experiments in which beams of light were passed through crystals. Nowadays a synthetic substance called polaroid is available for this demonstration (Fig. 10.20a). When light from a bulb passes through a single piece of polaroid P_1 its intensity is cut down to half. Rotating P_1 seems to have no effect on the transmitted beam since the transmitted intensity remains constant. Now let an identical piece



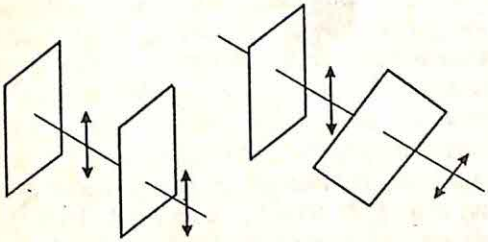


Figure 10.20: (a) Passage of light through two polaroids P_2 and P_1. The transmitted fraction falls from 1 to 0 as the angle between them varies from 0° to 90°. Notice that the light seen through a single polaroid P_1 does not vary with angle. (b) Behaviour of the electric vector when light passes through two polaroids. The transmitted polarisation is the component parallel to the polaroid axis. The double arrows show the oscillations of the electric vector.

polaroid P_2 be placed before P_1. As expected, the light from the bulb is reduced in intensity on passing through P_2 alone. But now rotating P_1 has a dramatic effect on the light coming from P_2! In one position, the intensity transmitted by P_2 followed by P_1 is nearly zero. When turned by 90° from this position, P_1 transmits nearly the full intensity emerging from P_2 (Fig. 10.20a). Clearly, the light transmitted by P_2 has some special property which the light from the bulb did not have.

This experiment can be understood using the properties of transverse waves. A plane monochromatic (single frequency) wave is described by giving the direction of travel, intensity and frequency. To give a *complete* description, we must specify the manner in which the electric field oscillates in the transverse plane ($x - y$ plane in Fig. 10.7). This property is called *polarisation.* For example, in Fig. 10.7 the tip of the electric vector oscillates back and forth in a straight line, namely, the x-axis. We call this wave *linearly polarised* along x. It is denoted by a double headed arrow since the electric vector points along both x and $-x$ at different times in a cycle. We can similarly think of a wave linearly polarised along the y direction or any other direction in the transverse plane. The space variation of the electric field in a linearly polarised wave is shown in Fig. 10.21a.

Now take a wave linearly polarised along a direction making an angle θ with the x-axis. We can resolve this into two waves. One has electric field $E\cos\theta$ and is linearly polarised along x and the other $E\sin\theta$ and is linearly polarised along y. We can now understand the experiment with two polaroids of Fig. 10.20. Let us assume that each polaroid has a special direction in its plane, which we call its *axis*. It is assumed that it transmits only the component of the electric field along the axis and the component perpendicular to the axis is absorbed. In Fig. 10.20b, the light transmitted by P_2 is polarised along its axis, as shown by the double arrows. The second polaroid P_1 transmits only a fraction $\cos\theta$ of the amplitude, polarised along *its* axis. The intensity transmitted is a fraction $\cos^2\theta$. As a function of the angle θ, this is a cosine curve with two maxima (at $\theta = 0°$ and 180°) per rotation. To see this we can write

$$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta). \qquad (10.33)$$

The two polaroids in an arrangement like Fig. 10.20 are called the polariser and analyser.

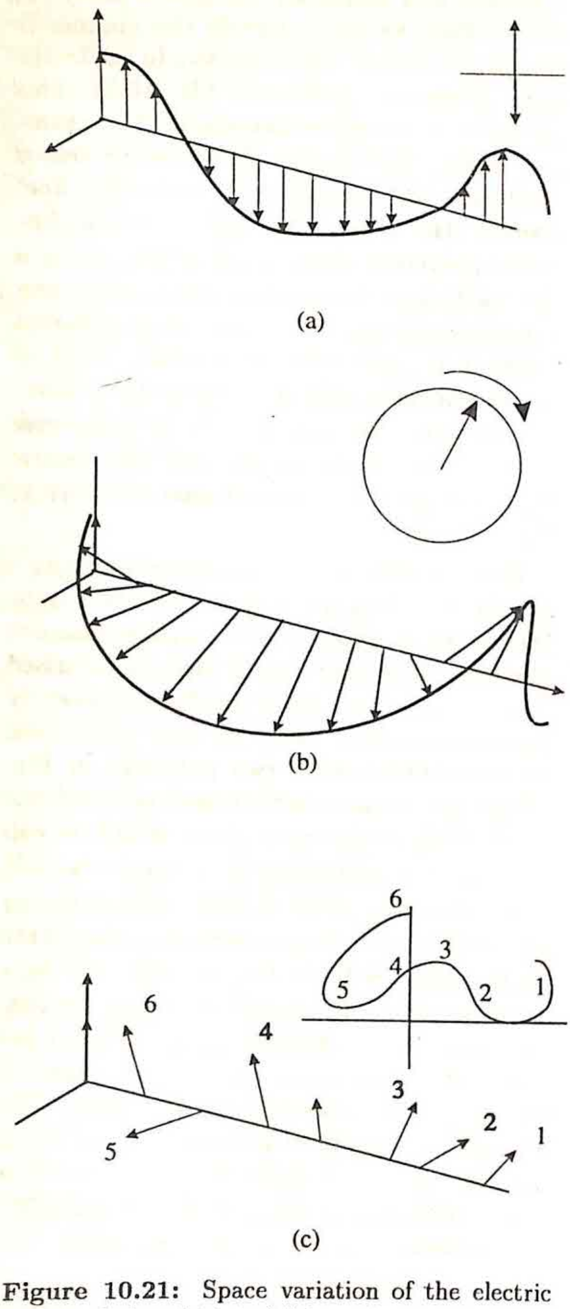


Figure 10.21: Space variation of the electric field for light which is (a) linearly polarised, (b) circularly polarised, and (c) unpolarised. Notice the random variation in the third case. The diagram on the right shows the path taken by the tip of the electric vector as time varies.

★ The basic reason for the action of the polaroid is that the molecules of a coloured substance (dye) point along a particular axis when the material is prepared. The movement of the electrons along the molecule allows them to absorb the radiation polarised in that direction. As a first approximation, polaroid transmits one component without absorption and absorbs the perpendicular component completely.

Two fairly common devices use polarised light. One is the liquid crystal display (LCD) found in many watches and calculators. Liquid crystals have long molecules whose direction can be controlled by applying electric fields. This is used to modify the light produced by a polariser so that its polarisation is perpendicular to the axis of an analyser which cuts it out. These dark regions can be controlled with applied voltages and used to form letters and numbers.

Some sunglasses (dark glasses) have polaroids. Glare from sunlight reflected from water or some horizontal surface can be reduced if the polaroid cuts out horizontally polarised light. (See next section). ★

We now come to the behaviour of light from a bulb or the sun passing through a single polaroid (Fig. 10.20). Half the light from the bulb is transmitted for *all* positions of the polaroid. This is because the electric field of the incident light does not have any fixed direction. We have already remarked that the radiation from a source like a bulb is produced by electrons in motion. We expect the direction of the electric field in the transverse plane to change in a more or less irregular way because of the irregular motion of the radiating charges. Light of this kind is called *unpolarised*. The typical way in which the field might vary in space is sketched in Fig. 10.21c. The average transmitted intensity remains constant as we observe unpolarised light through a polaroid which is rotated. The average of Eq.(10.33)

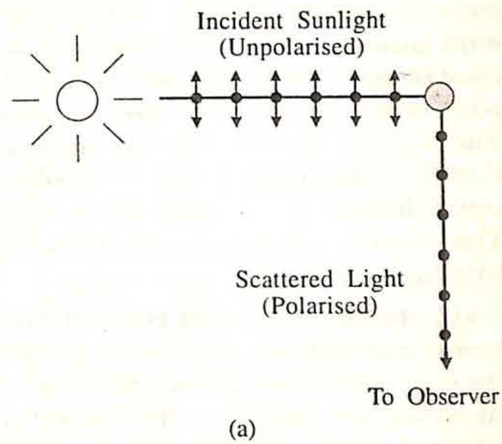
for the transmitted intensity, when we allow θ to vary over all angles (0 to 2π) equals 1/2, since the average of $\cos 2\theta$ is zero.

★ Light linearly polarised at 45° to the x-axis can be resolved into equal components along x and y axis which are in phase. What happens if we change the relative phase of the x and y components by 90°? (This can be done using a thin slice of crystal which has different refractive indices for electric fields along x and y. Because of the 90° phase difference it is called a quarter wave plate). Two simple harmonic motions with equal amplitude and 90° phase difference at right angles to each other combine to give circular motion. (Chapter 12, class XI text book). The tip of the electric vector describes a circle. We refer to such a wave as *circularly polarised*. The space variation of the electric field is shown in Fig. 10.21b. In the most general way of combining simple harmonic motions, we get motion in an ellipse. This produces *elliptic* polarisation.

Notice that circularly polarised light also gives constant average intensity when viewed through a polaroid at any angle. How to distinguish this from unpolarised light? The answer is that the same quarter wave plate can be used to remove the 90° phase difference between E_x and E_y. We get back linear polarisation which can then be detected by rotating an analyser. The same experiment with unpolarised light gives a different result. There is no stable phase relationship between E_x and E_y to start with. Adding an extra 90° of phase to one therefore makes no difference. The light remains unpolarised after passing through the quarter wave plate. The analyser which follows therefore shows no variation in intensity as it is rotated. ★

10.8.2 Polarisation by scattering and reflection

The light from a clear blue portion of the sky shows a rise and fall of intensity when viewed through a polaroid which is rotated. This is nothing but sunlight which has changed its direction (having been scattered) on encountering the molecules of the earth's atmosphere. As Fig. 10.22a shows, the incident sunlight is unpolarised. The dots stand for polarisation perpendicular to the plane of



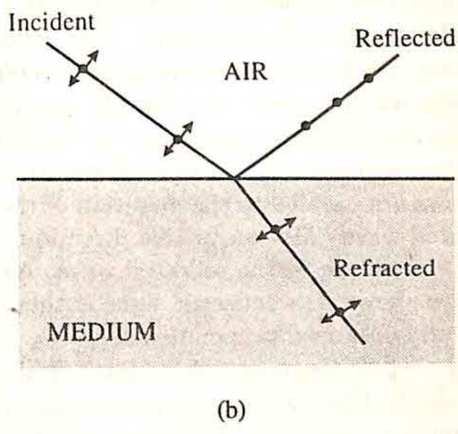


Figure 10.22: (a) Polarisation of the blue scattered light from the sky. The incident sunlight is unpolarised (dots and arrows). A typical molecule is shown. It scatters light by 90° polarised normal to the plane of the paper (dots only). (b) Polarisation of light reflected from a transparent medium at the Brewster angle (reflected ray perpendicular to refracted ray).

the figure. The double arrows show polarisation in the plane of the figure. (There is no phase relation between these two in unpolarised light). Under the influence of the electric field of the incident wave the electrons in the molecules acquire components of motion in both these directions. We have drawn an observer looking at 90° to the direction of the sun. Clearly, charges accelerating parallel to the double arrows do not radiate energy towards this observer since their acceleration has no transverse component. The radiation scattered by the molecule is therefore represented by dots. It is polarised perpendicular to the plane of the figure. This explains the polarisation of scattered light from the sky.

Fig. 10.22b shows light reflected from a transparent medium, say water. As before the dots and arrows indicate that both polarisations are present in the incident and refracted waves. We have drawn a situation in which the reflected wave travels at right angles to the refracted wave. The reflected wave is produced by the oscillating electrons in the water. These move in the two directions transverse to the radiation from wave in the medium, i.e., the *refracted* wave. The arrows are parallel to the direction of the *reflected* wave. Motion in this direction does not contribute to the reflected wave. As the figure shows, the reflected light is therefore linearly polarised perpendicular to the plane of the figure (represented by dots). This can be checked by looking at the reflected light through an analyser. The transmitted intensity will be zero when the axis of the analyser is in the plane of the figure, i.e., the plane of incidence.

Example 10.8: What should be the angle of incidence in Fig. 10.22b so that the reflected and refracted rays are perpendicular?

Answer: From the geometry of the figure

$$i + r = 90°$$

The refractive index n

$$= \frac{\sin i}{\sin r}$$

$$= \frac{\sin i}{\sin(90° - i)} = \tan i.$$

The special angle of incidence satisfying this condition is called the Brewster angle. For $n = 1.5$ it is approximately 57°.

★ For simplicity we have discussed scattering by 90° and reflection at the Brewster angle. In these cases only one of the two polarisations is present. An other angles, both are present but one is stronger than the other. There is no stable phase relationship between the two perpendicular components since these are derived from two perpendicular components of an unpolarised beam. When such light is viewed through a rotating analyser, one sees a maximum and a minimum of intensity but not complete darkness. This kind of light is called *partially polarised.* ★

The scattering of light by molecules was intensively investigated by C.V.Raman and his collaborators in Calcutta in the nineteen twenties. Raman was awarded the Nobel Prize for Physics in 1930 for this work.

10.8.3 Rotation of the plane of polarisation

When linearly polarised light travels through glass or water, the direction of linear polarisation remains fixed. However, some substances like sugar solution show the remarkable property of rotating this direction in the

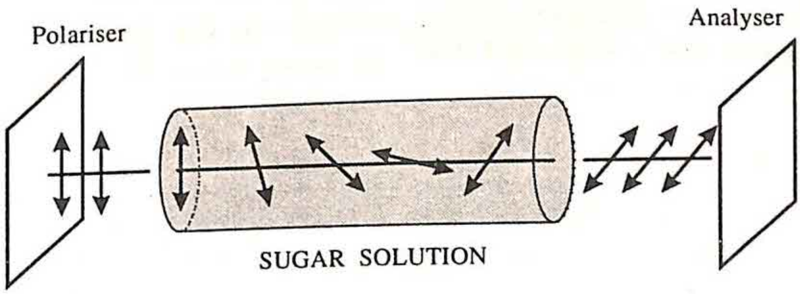


Figure 10.23: Linearly polarised light from the first polaroid enters sugar solution in a cylindrical tube. The direction of polarisation rotates. This is detected by a second polaroid which gives maximum intensity only after turning it through an angle.

transverse plane. This effect is called *optical rotation* or *optical activity*. Fig. 10.23 illustrates an experiment to measure this rotation. A polariser produces linearly polarised light which travels along a tube of sugar solution. The polarisation of the emerging beam can be found using an analyser. The angle of rotation is found to be proportional to the length traversed and to the concentration of sugar.

The pattern of polarisation vectors in Fig. 10.23 is like a left handed screw. As you know, such an object is not identical to its mirror image (which is a right handed screw). One can ask why the polarisation of light in sugar solution rotates like a left handed screw rather than a right handed screw. The answer is that a sugar molecule is not identical to its mirror image. Chemists can prepare a molecule which is the mirror image of the naturally occuring sugar. In solution, this rotates the plane of polarisation in a direction opposite to that shown in Fig. 10.23. It is interesting that in living organisms most of the molecules occur in only one of the two mirror image forms. (Sugar comes from the sugarcane plant!) A chemical preparation starting from substances not showing optical activity gives an equal mixture of the two mirror image molecules. This is called a *racemic* mixture. However, when crystals form from such a mixture, the two mirror image molecules sometimes get separated. This was how the French chemist (and biologist) Pasteur discovered mirror image molecules and their opposite optical rotation. Chemists call such a pair *enantiomers* and the two forms are called dextro and laevo (which mean right and left and are abbreviated by d and l) depending on the sign of optical activity.

A very simple molecule like nitrogen, N_2 consists of two atoms joined by a chemical bond. It looks identical to its mirror image. In a gas or in solution, these molecules occur in all possible orientations. This entire assembly looks identical to its mirror image. It cannot produce optical rotation. Only molecules which are not identical to their mirror images can produce this effect.

★ In the nineteen seventies, very delicate experiments revealed that even a gas of *atoms* produces minute optical rotation. This suggests that even atoms are not strictly identical to their mirror images, though the effect is very weak. This is

usually neglected for all purposes in atomic and molecular physics. However, such weak forces are of great importance for particle physics. The optical rotation produced by atoms, tiny as it is, was an important confirmation of the theory for which Glashow and Weinberg (US) and Salam (Pakistan) won the Nobel prize in 1979. ★

Summary

1. A *wavefront* is the locus of points having the same phase of oscillation. *Rays are lines perpendivular to the wavefront,* which show the direction of propagation of energy. *The time taken for light to travel from one wavefront to another is the same along any ray.*

2. *Huygens' construction* is based on the principle that every point of a wavefront is a source of secondary wavefront. The envelop of these wavefronts i.e. the surface tangent to all the secondary wavefront gives the new wavefront.

3. The law of reflection ($i = r$) and the Snell's law of refraction

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2} = \frac{n_2}{n_1} = n_{21}$$

can be derived using the wavetheory. (Here v_1 and v_2 are the speed of light in media 1 and 2 with refractive index n_1 and n_2, respectively). The derivation uses the principles underlined in 1 above. The frequency ν remains the same as light travels from one medium to another. The speed v of a wave is given by

$$v = \lambda / T = \lambda \nu$$

where λ is the wavelength of the wave, and T is the period of oscillation.

4. Emission, absorption and scattering are three processes by which matter interacts with radiation. In emission, an accelerated charge radiates and loses energy; in absorption, the charge gains energy at the expense of the electromagnetic wave; in scattering,the charge accelerated by incident electromagnetic wave radiates in all direction.

5. Two sources are *coherent* if they have the same frequency and a stable phase difference. In this case,the total intensity I is *not* just the sum of individual intensities I_1 and I_2 due to the two sources but includes an interference term:

$$I = I_1 + I_2 + 2k\mathbf{E}_1 \cdot \mathbf{E}_2$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric fields at a point due to the sources.
The interference term averaged over many cycles is zero if (a) the sources have different feequencies or (b) the sources have the same frequency but no stable phase difference. For such incoherent sources, $I = I_1 + I_2$.

6. *In Young's experiment,* two parallel and very close slits S_1 and S_2 (illuminated by another narrow slit) behave like two coherent sources and produce on a screen a pattern of dark and bright bands - interference fringes. For a point P on the screen,

$$S_2P - S_1P \simeq \frac{y_1 d}{D_1}$$

where d is the separation between the two slits, D_1 is the distance between the slits and the screen and y_1 is the distance of the point P from the central fringe. For constructive interference (bright band), the path diffference must be an integer multiple of λ i.e.

$$\frac{y_1 d}{D_1} = n\lambda \text{ or } y_1 = n\frac{D_1 \lambda}{d}$$

The separation Δy_1 between adjacent bright (or dark) fringes is,

$$\Delta y_1 = \frac{D_1 \lambda}{d}$$

using which λ can be measured.

7. *The colours shown by thin films* are due to interference between two beams, one reflected from the top surface of the film and the other from the bottom. The path difference between the two may give constructive interference for one colour and destructive interference for another. Hence the reflected light is coloured.

8. *Diffraction* refers to light spreading out from narrow holes and slits, and bending around corners and obstacles. The single-silt diffraction pattern shows the central maximum ($\theta = 0$), zero intensity at angular separation $\theta = \pm n\lambda \quad (n \neq 0)$ and secondary maxima at $\theta = \pm(n + 1/2)\lambda$ $(n \neq 0)$. Different parts of the wavefront at the slit act as secondary sources; diffraction pattern is the result of interference of waves from these sources.

   An aperture of size a sends diffracted light into an angle $\approx \lambda/a$. Ray-optics is a good approximation for a beam up to a distance a^2/λ.

9. *Doppler effect* is the shift in frequency of light when there is a relative motion between the source and the observer. It is given by

$$\frac{\Delta \nu}{\nu} \approx \frac{v_r}{c} \text{ for } \frac{v}{c} << 1$$

   where v_r is the radial component of the relative velocity $\mathbf{v}$. The effect can be used to measure the speed of an approaching or receding object. Doppler effect occurs for sound also - it is a characteristic property of waves.

10. *Polarization* specifies the manner in which electric field $\mathbf{E}$ oscillates in the plane transverse to the direction of propagation of light. If $\mathbf{E}$ oscillates back and forth in a straight line, the wave is said to be linearly polarized. If the direction of $\mathbf{E}$ changes irregularly, the wave is unpolarized.

    When light passes through a single polaroid P_1, light intensity is cut to half, independent of the orientation of P_1. When a second polaroid P_2 is interposed between the source and P_1, light emerging out of P_2 is not at all transmitted by P_1 at one orientation, and is transmitted fully when P_1 is turned 90^o from that orientation. This happens

because the transmitted polarization by a polaroid is the component of $\mathbf{E}$ parallel to its axis.

Unpolarized sunlight scattered by the atmosphere or reflected from a medium gets (partially) polarized.

11. *Linearly polarized* light passing through some substances like sugar solution undergoes a rotation of its direction of polarization, proportional to the length transversed and the concentration of the substance. This effect is known as *optical activity*.

Exercises

10.1 Explain the term 'wavefront'. Describe Huygens' construction for the propagation of wavefronts in a medium.

10.2 What is the geometrical shape of the wavefront in each of the following cases:

(a) light diverging form a point source.

(b) light emerging out of a convex lens when a point source is placed at its focus.

(c) the portion of the wavefront of light from a distant star intercepted by the earth.

10.3 Derive the law of reflection from a plane surface using the wave theory of light.

10.4 Light of wavelength 5000 Å falls on a plane reflecing surface. What are the wavelength and frequency of the relfected light? For what angle of incidence is the reflected ray normal to the incident ray?

10.5 Derive Snell's law of refraction using the wavetheory of light.

10.6 (a) The refractive index of glass is 1.5. What is the speed of light in glass? (Speed of light in vacuum is $3.0 \times 10^8\ ms^{-1}$)

(b) Is the speed of light in glass independent of the colour of light? If not, which of the two colours red and violet travels slower in a glass prism?

10.7 Monochromatic light of wavelength 600 nm is incident form air on a glass surface. What are the wavelength, frequency and speed of the refracted light? Refractive index of glass is 1.5

10.8 Explain the statement: "To get interference we need the same frequency and a stable phase difference between the sources".

10.9 A region is illuminated by two sources of light. The intensity I at each point is found to be equal to $I_1 + I_2$, where I_1 is the intensity of light at the point when source 2 is absent. I_2 is similarly defined. Are the sources coherent or incoherent Explain.

10.10 In a Young's double-slit experiment, the slits are seperated by 0.28 mm and the screen is placed 1.4 m away. The distance between the central bright fringe and the fourth bright fringe is measured to be 1.2 cm. Determine the wavelength of light used in the experiment.

10.11 What is the Brewster angle for air to glass transition? (refractive index of glass = 1.5)

10.12 Two polaroids are placed 90^o to each other and the transmitted intensity is zero. What happens when one more polaroid is placed between these two bisecting the angle between them?

10.13 A diffraction grating 1 cm wide has 1200 lines and is used in second order. What is the diffraction angle for light of wavelength 550 nm?

10.14 Estimate the distance for which ray optics is good approximation for an aperature of 4 mm and wavelength 400 nm.

10.15 Two towers on top of two hills are 40 km apart. The line joining them passes 50 m above a hill halfway between the towers. What is the longest wavelength of radiowaves which can be sent between the towers without appreaciable diffraction effects?

10.16 The 6563 Å H_α line emitted by hydrogen in a star is found to be redshifted by 15 Å. Estimate the speed with which the star is receding from the earth.

Additional Exercises

10.17 Monochromatic light of wavelength 589 nm is incident from air on a water surface. What are the wavelength, frequency and speed of a) reflected, and (b) refracted light? Refractive index of water is 1.33.

10.18 Explain how Newton's corpuscular theory predicts the speed of light in a medium, say, water, to be greater than the speed of light in vacuum. Is the prediction confirmed by experimental determination of the speed of light in water? If not, which alternative picture of light is consistent with experiment?

10.19 You have learnt in the text how Huygens' principle leads to the laws of reflection and refraction. Use the same principle to deduce directly that a point object placed in a front of a plane mirror produces a virtual image whose distance from the mirror is equal to the object distance from the mirror.

10.20 Let us list some of the factors which could possibly influence the speed of wave propagation:

(i) natute of the source
(ii) direction of propagation
(iii) motion of the source and/or observer
(iv) wavelength
(v) intensity of the wave

On which of these factors, if any, does

(a) the speed of light in vacuum
(b) The speed of light in a medium (say, glass or water)
depend?

10.21 For sound waves, the Doppler formula for frequency shift differs slightly between the two situations: (i) source at rest-observer moving, and (ii) source moving-observer at rest. The exact Doppler formulas for the case of light waves in vacuum are, however, strictly indentical for these situations. Explain why this should be so. Would you expect the formulas to be strictly identical for the two situations in case of light travelling in a medium?

10.22 Answer the following questions:

(a) When monochromatic light is incident on a surface separating two media, the reflected and refracted light both have the same frequency as the incident frequency. Explain why?

(b) When light travels from a rarer to a denser medium, it loses some speed. Does the reduction in speed imply a reduction in the energy carried by the light wave?

(c) A narrow pulse of light is sent through a medium. Will you expect the pulse to retain its shape as it travels through the medium?

(d) In the wave picture of light, intensity of light is determined by the square of the amplitude of the wave. What determines the intensity of light in the photon picture of light

(e) The speed of light in still water is c/n, where n is the refractive index of the water. What is the speed of light in a stream of water flowing at a steady speed of v relative to the observer?

10.23 What is the effect on the interference fringes in a Young's double-slit experiment due to each of the following operations:

(a) the screen is moved away from the plane of the slits;

(b) the (monochromatic) source is replaced by another (monochromatic) source of shorter wavelength;

(c) the separation between the two slits is increased;

(d) the source slit is moved closer to the double-slit plane;

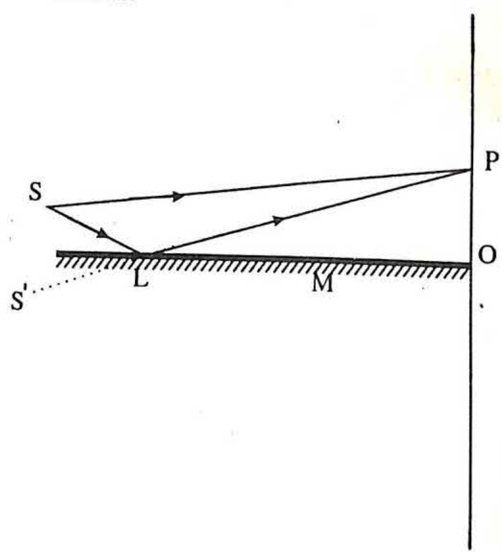
(e) the width of the source slit is increased;

(f) the widths of two slits are increased;

(g) the monochromatic source is replaced by source of white light?

(In each operation, take all parameters, other than the one specified, to remain unchanged).

10.24 Figure shows an outline of Lloyd's mirror experiment. M is a plane mirror; S is a narrow slit illuminated by some source of light (not shown) and S' is the image of S in M. M, S and S' are in in a plane prependicular to the paper. O is the line of intersection of the mirror and the screen.

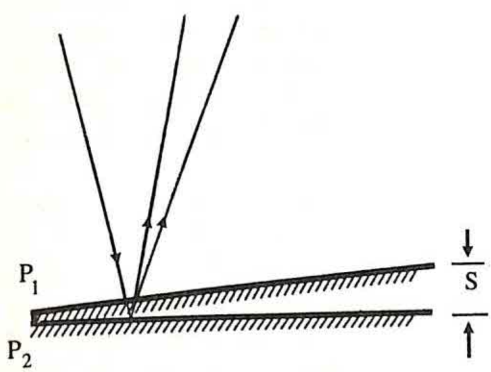


(a) What is the origin of the fringes observed on the screen?

(b) Why is the slit S placed so as to have very oblique angle of incidence of light striking the mirror?

(c) The two path length PS and PS' are *equal* when P coincides with O. Yet the fringe at O is found in the experiment to be dark not bright.
What does this observation imply?

10.25 Figure shows two flat glass plates P_1, and P_2 placed nearly (but not exactly) parallel forming an air wedge. The plates are illuminated normally by monochromatic light and viewed from above. Light waves reflected from the upper and lower surfaces of the air wedge give rise to an interference pattern:



(a) Show that the separation between two successive bright (or dark) fringes is given by $\frac{\lambda \ell}{2S}$ where ℓ is the length of each plate and S is the separation between the plates at the open end of the wedge.

(b) In the experiment, a *dark* fringe is observed along the line joining the two plates. Why?

(c) If the space between the glass plates is filled with water, what changes in the fringe pattern do you except to see, if at all?

(d) Suggest a way of obtaining a bright fringe along the line of contact of the two plates in this experiment.

10.26 Give the shape of interference fringes observed

(a) in a Young's double-slit experiment

(b) in the air wedge experiment

(c) in the Lloyd's mirror experiment

(d) when a small lamp is placed before a thin mica sheet and lightwaves reflected from the front and back surfaces of the sheet combine to produce interference pattern on a screen behind the lamp. (Pohl's experiment)

(e) from a thin air film formed by placing a convex lens on top of a flat glass plate (Newton's arrangement).

10.27 (a) Red light of wavelength 6500 Å from a distant source falls on a slit 0.50 mm wide. What is the distance between the two dark bands on each side of the central bright band of the diffraction pattern observed on a screen placed 1.8 m from the slit?

(b) What is the answer to (a) if the slit is replaced by a small circular hole of diameter 0.50 mm ?

10.28 Answer the following questions:

(a) In a single-slit diffraction experiment, the width of the slit is made double the original width. How does this affect the size and intensity of the central diffraction band?

(b) In what way is diffraction from each slit related to the interference pattern in a double-slit experiment?

(c) When a tiny circular obstacle is placed in the path of light from a distant source, a brightspot is seen at the centre of the shadow of the obstacle. Explain why?

(d) Two students are seperated by a 7 m partition wall in a room 10 m high. If both light and sound waves can bend around obstacles, how is it that the students are unable to see each other even though they can converse easily.

(e) Geometrical optics is based on the assumption that light travels in a straight line. Diffraction effects (observed when light propagates through small apertures/slits or around small obstacles) disprove this assumption. Yet the geometrical optics assumption is so commonly used in understanding location and several other properties of images in optical instruments. What is the justification?

10.29 Answer the following question:

(a) When a low-flying aircraft passes overhead, we sometimes notice a slight shaking of the picture on our TV screen. Suggest a possible explanation.

(b) Thin flims such as a soap bubble or a thin layer of oil on water show beautiful colours when illuminated by white light. Explain the observation.

(c) In a thin-film interference experiment (the experiment on Newton's rings, for example) the cental fringe of the pattern is dark when viewed by reflected light, and bright when viewed by transmitted light. Why?

(d) If white light is used in the air wedge interference experiment (see exercise 10.25) or the Newton's rings experiment, the colour observed in the reflected light is complementary to that observed in the light transmitted through the same point. Why?

(e) As you have learnt in the text, the principle of linear superposition of wave displacement is basic to understanding intensity distributions in diffraction and interference patterns. What is the justification of this principle?

10.30 At a given point in space, circularly polarized light produces equal amplitude vibrations along x and y with a 90^o phase difference. $E_x \doteq E_0 \cos \omega t$ and $E_y = E_0 \sin \omega t$. Let x' and y' be a new set of axes rotated by θ in the $x - y$ plane. If

the same vibrations $E_0 \cos\omega t$ and $E_0 \sin\omega t$ are present along x' and y', show that the result is still circularly polarized light with a different phase. Show that if E_y is changed in phase by π, the circle is traversed in the opposite sense.

10.31 Show that the two oppositely circularly polarised beams of the same frequency and equal amplitude combine to give linear plarisation. What should one do to the relative phase of the two beams to rotate the direction of linear polarisation? Can you use this to understand what happens to the two opposite circular polarisation in sugar solution?

10.32 Two polaroids are placed at 90^o to each other and the transmitted intensity is zero

(a) what happens when one more polaroid is placed between these two bisecting the angle between them?

(b) (N-1) more polaroids are inserted between two crossed polaroids (at 90^o each other). Their axes are equally spaced. How does the transmitted intensity behave for large N? [**Hint:** Calculate a few special cases e.g $N = 4, 8, \cdots$]

10.33 A halfwave plate is a device which introduces a phase difference of π between E_x and E_y. What is its effect on

(a) lineraly polarised light making angle θ to the x-axis

(b) circularly polarised light?

10.34 Sodium light has two wavelengths $\lambda_1 = 589$ nm and $\lambda_2 = 589.6$ nm. As the path difference increases, when is the visibility of the frings a minimum?

10.35 In deriving the single slit diffraction pattern, it was stated that the intensity is zero at angles of $n\lambda/a$. Justify this by suitably dividing the slit to bring out the cancellation.

10.36 In a pinhole camera, a box of length L has a hole of a radius a in one wall. When the hole is illuminated by a parallel beam, the size of spot of light is large. Show that it is also very large when a is small, due to diffraction. Assume that the spread due to diffraction just adds to the geometrical spread and find the minimum size of the spot.

10.37 Two coherent beams intersect at a small angle θ. What is the spacing of the interference fringes on a screen whose normal bisects the directions of two beams? Instead of a screen, a photographic film is used. When it is developed, the fringes appear as opaque and transparent regions. The film is then used as a grating. What happens when one of the two beams which produced the interference is allowed to fall on this grating?

Christiaan Huygens (1629-1695) Dutch physicist, astronomer, mathematician and the founder of the wave theory of light. His book, "Treatise on light", makes fascinating reading even today. He brilliantly explained the double refraction shown by the mineral calcite in this work in addition to reflection and refraction. He was the first to analyse circular and simple harmonic motion and designed and built improved clocks and telescopes. He discovered the true geometry of Saturn's rings.

Jean - Augustin Fresnel (1788-1827) French engineer who took to the study of wave optics. Although Young had demonstrated two slit interference, Fresnel was the first to explain the whole range of diffraction effects from all apertures and at all distances. A dramatic prediction of his theory (soon verified) was that at the very centre of the shadow of a circular obstacle, there should be a bright spot. His further work includes formulae for reflection and transmission by a plane surface which are still known by his name, the general theory of light propagation in crystals, the motion of light in a moving medium, and much more.

Jean-Bernard Foucault (1819-1868) French experimenter whose measurement of the speed of light in water conclusively proved the wave theory. By his rotating mirror method, he was able to make accurate measurement is the laboratory. His other well knwon experiment is the demonstrative of the earth's rotation by the change in the plane of oscillation of a pendulum. His study of gyroscopes ultimately led to modern devices which are among the most accurate methods of guiding a moving plane, ship or rocket.

CHAPTER **11**

Ray Optics and Optical Instruments

11.1 Introduction

Much of what we know about the world around us is through light and the sense of vision. We have seen in the last two Chapters that light is an electromagnetic wave (Chapter 9), but that its wavelength (4×10^{-7}m to 7×10^{-7} m) is short enough so that bending and spreading of light waves (diffraction) by obstracles can often be neglected. Also, if the sources are incoherent, interference of light does not occur (Chapter 10). Under these conditions, light travels (propagates) from one point to another in a straight line joining them called a ray and intensities can be added to give total intensity. These rays can be, reflected by mirrors, refracted by transparent media such as glass or water, focussed by lenses and mirrors, etc. In this Chapter, we study the behaviour of light rays under different conditions, and various optical instruments including the eye, for forming images.

We begin by briefly describing sources of light, and units in which intensity of light is

measured (Section 11.2). We make a brief excursion into the ways in which the speed of light, a very large and fundamentally important quantity, has been measured (Section 11.3). We then consider the reflection of light from spherical mirrors (Section 11.4). The next several sections are concerned with the refraction of light by transparent slabs, prisms and lenses (Section 11.5 and 11.6).

The refractive index of a medium depends on the wavelength of the light going through it. This spread of refractive index, called dispersion (Section 11.7) has many consequences, some colourful. The spectrum of colours, different types of spectra and instruments for looking at or analysing spectra are the subject of this section. Some natural optical phenomena e.g. the colour of the sky, and the rainbow, are described and explained in the next section (Section 11.8). The dependence of refractive index on wavelength, as well as several simplifications we have made in our description of refracting lenses and reflecting mirrors, all lead to imperfections or defects in images formed by them. These are mentioned in Section 11.9. Finally, we try to understand the working of different optical instruments, starting with the eye, and including the camera, the binocular, the microscope and the telescope (Section 11.10). The resolving power of optical instruments is also briefly analyzed in this section.

11.2 Sources of light, luminosity and photometry

11.2.1 Types of light sources

From common experience, we know of many sources of light. The sun, a carbon arc, an electric bulb, and the fluorescent tube are some of the familiar sources of light. Light sources can be broadly classified into three categories on the basis of the processes involved in the emission of radiation by them: thermal, gas discharge and luminescent. Light from a thermal source such as an incandescent bulb which emits light simply because it is hot (see Section 11.8c, class XI Physics textbook) contains a whole continuous range of visible wavelengths. Light from a gas discharge tube contains a few wavelength bands which give it a characteristic colour depending on the nature of the gas, etc. For example a neon tube light is red. In a fluorescent tube light, the gas discharge produces partly visible light and partly ultraviolet radiation. This ultraviolet radiation acts on material called phosphors coated on the glass. Then there is absorption of the visible and ultraviolet radiation, and reemission of visible radiation. The fluorescent substances in the tube light (called phosphors) commonly used are mixtures of calcium tungstate, zinc silicate and cadmium borate. The phenomenon we have described is called photoluminescence (emission of light after absorbing some electromagnetic radiation). Luminescence can also be induced by other methods for example, electrically as in light emitting diodes. This is called electroluminescence. The firefly produces light by means of complex chemical reactions (chemiluminescence).

11.2.2 Luminous intensity, flux, and illuminance

Since light is an electromagnetic wave, the intensity of light passing through a given area is connected with the energy(of the electric and magnetic fields of the em wave) flowing in a unit time through that area. The natural unit of the intensity of light is energy per unit time, or power (watt in SI units). Now this flux of light energy, falling on the eye, excites the sensation of vision.

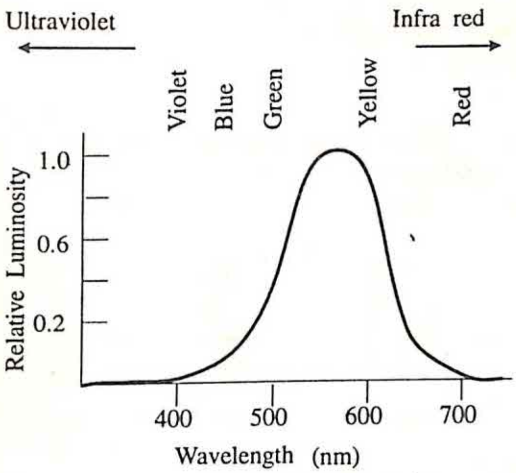


Figure 11.1: Standard luminosity curve. (Note that the 'colours' visible to the human eye are from violet to red and of these, the green and yellow colours are prominently seen).

So there is, naturally, a unit of luminous power related to perception by the human eye. This is called the lumen. We define it below. There are two related definitions, of luminous intensity and illuminance, which are also given below.

Before going into these, we emphasize the fact that the sensitivity of the eye to light of various wavelengths in the visible region is different. Fig 11.1 shows the relative sensitivity of the eye for light in the visible region. We see that the sensitivity peaks at about 555 nm (or 5.5×10^{-7} m; 5500 Å). This is in the yellow green region. We notice that the sensitivity decreases rather rapidly for both longer and shorter wavelengths and becomes small indeed for $\lambda > 700$ nm and for $\lambda < 400$ nm. Sensitivity here means the sensation of sight for a given power per unit area of light of that wavelength incident on the eye. For the same power per unit area, yellow light produces the strongest sensation of sight.

The unit of *luminous* power is defined with respect to the perception by the eye. Since the eye has varying sensitivity for different wavelengths, we choose one wavelength, namely the peak sensitivity wavelength of 555 nm (or about 5.4×10^{14} Hz frequency). The unit of luminous power, called *lumen* is (1/683) of a watt of light (electromagnetic wave) at this wavelength (683 lumens at 555 nm wavelength make one watt). Clearly, at other wavelengths, longer or shorter, one watt will appear less luminous, i.e. will have a smaller number of lumens. The abbreviated symbol for lumen is lm.

Why 683 lumens? The answer to this question has to do with history. Originally, the light from a certain type of candle was chosen as the standard of light power. Now better standards are available (such as monochromatic sources with a given wattage!). However in relation to the luminous power from the old standard candle of one lumen, the watt is 683 lumens.

A light source often radiates in all directions. For example a point source radiates *equally* in all directions. What matters often is the amount of light reaching a given *solid angle* (such as that subtended by the eye, or a mirror) or a given *area*. So there are two other definitions. One is of *luminous intensity* or luminous power per unit solid angle. This has the unit *candela* (abbreviated cd). This is the SI unit of luminous intensity, one of the SI base units (see Appendix A of Part I of Class XII Physics textbooks). It is one lumen per steradian[1] (Or since cd is the base unit, one should say that a lumen is one candela steradian). The relation between a

[1]The steradian, abbreviated as sr, is the solid angle obtained as follows:- Consider a sphere of radius r. On the surface of the sphere (total area $4\pi r^2$) cut off an area equal to r^2. This area forms a solid angle equal to one steradian at the centre of the sphere. The total solid angle of the centre is clearly 4π.

lumen and a candela is simple. Suppose we have a point source of luminous intensity one candela or one lumen/steradian. Over the total solid angle of 4π, the luminous power radiated is 4π lumens. Both candela and lumen have the physical dimensions of power.

Finally, illuminance is the name for luminous power per *unit area*. It unit is *lux*, which is *one* lumen per square meter. For example, if we have a point source of luminous power Φ lumens, the power is uniformly distributed over a spherical surface of area $A = 4\pi r^2$ at a distance r from the source. Thus the illuminance E is

$$E = \frac{\Phi}{A} = \frac{\Phi}{4\pi r^2}\ \mathrm{lm/m^2} = \frac{\Phi}{4\pi r^2}\ \mathrm{lux} \quad (11.1)$$

Now in terms of the luminous intensity I, a source of luminous power Φ lm radiates $(\Phi/4\pi)$ cd per steradian, or

$$I = \frac{\Phi}{4\pi} \quad (11.2)$$

therefore

$$E = \frac{4\pi I}{4\pi r^2} = \frac{I}{r^2}\ \mathrm{lm/m^2} \quad (11.3)$$

where I is the source intensity in candelas.

We see as expected that the illuminance falls off as $(1/r^2)$.

Light sources emit, in the visible region, only a certain fraction of the energy required for powering them. For example, a 40 W tungsten lamp produces a luminous power of only about 465 lumens! (Small indeed if *one* watt is 683 lumens at the peak sensitivity wavelength of 555 nm). We show for comparison the efficiencies of some normal light sources in Table 11.1. The efficiencies are in lm/W, a dimensionless unit. We notice that indeed the fluorescent lamp is four times as efficient in providing visible light as the 40 W tungsten lamp.

Table 11.1: Efficiencies of some common light sources.

Source (tungsten lamp)	Luminous Power (lumens)	Efficiency (lm/W)
40 W	465	~ 12
60 W	835	~ 14
500 W	9950	~ 20
30 W	1500	50

11.3 Velocity of light

The speed with which light travels is one of the most important quantities in physics. For a long time it was felt that light takes *no* time to go from one place to another, namely that it propagates instantaneously. Galileo was the first to suggest that light takes a finite time to travel between two points in space, and tried to measure the velocity of light (commonly given the symbol c). He and a partner stood on two hilltops about a kilometre apart. Each had a lantern and a shutter to cover it, and a clock. One partner would uncover his lantern and start the clock. The second partner would uncover *his* lantern on seeing the flash. The first partner would stop the clock as soon as the light from the second partner's lantern reached his eye. Thus the time recorded by the clock is the time it takes for light to go from one hilltop to the other, and back. This experiment failed because the speed of light is so large that the time taken for light to traverse twice the distance between the two hilltops is much less than the human response time which fluctuates widely.

The first indication of the magnitude of the velocity of light came from the observations of the astronomer Ole Römer in the

period 1667-1676. He noticed that the observed time interval between two successive eclipses of the moons of Jupiter depended on the time of the year. This time is found to be larger if the Earth is moving away from Jupiter and smaller if the Earth is moving towards it. He correctly argued that this is because the velocity of light is finite. The actual time taken for a satellite or moon of Jupiter to go around Jupiter once (period of revolution of the satellite) is the same. However, in this time, the earth might have moved a little away from Jupiter (position B in Fig 11.2) or a little toward it (position D). In the former case, we have to add to the true period T_0, the time taken for light to cover the distance which the earth has moved away in time T_0. If the earth is moving towards Jupiter, the observed period is less by nearly the same amount.

For example, the satellite Io takes nearly 42.5 hours to go around Jupiter, and this time varies by about 15 seconds between one eclipse and the next, depending on the earth's position. Since this time is small, Romer proposed the following method to find accurately the speed of light.

The duration between successive eclipses is measured when the earth is at point A (Fig. 11.2), nearest to Jupiter, so that the distance between the earth and Jupiter hardly changes from one of eclipse to the next eclipse. This is the true period of revolution of the satellite. Now suppose we start counting time (starting a clock) when the eclipse occurs at A. Knowing the true period, one can compute the number of eclipses in six months, to the nearest integer, and thus predict the time at which the eclipse will occur when the earth is near C (six months from A).

Romer found that eclipse actually occurred about 20 minutes later! Why is this? Clearly because light has to cover an extra distance AC as the earth (or the observer) moves from A (first eclipse) to C (last eclipse). As the earth moves away from Jupiter (which moves so little in six months that it can be assumed fixed; corrections for its motion can be made) we see that each successive eclipse will occur at a slightly later time, because light takes time to travel the extra distance. The delays add up from A to C; the total delay is the time taken for light to travel from A to C. (The correct value is nearly 16.6 minutes, rather than the 20 minutes Römer measured). Knowing the radius of the earth's orbit around the sun ($\sim 1.4 \times 10^{11}$ m), the velocity of light can be calculated. Römer obtained a value of 2.25×10^8 m/s. (The modern, correct value is close to 3×10^8 m/s). The speed of light is astonishingly large.

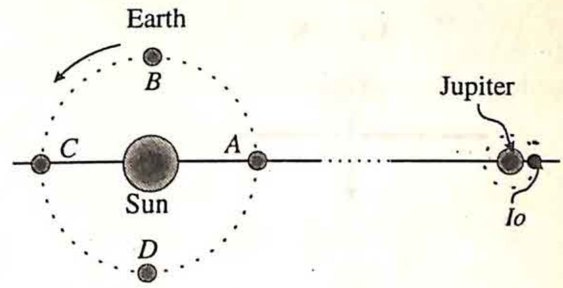


Figure 11.2: Römer's method of measuring the velocity of light. The earth going round the sun, the planet Jupiter, and its satellite Io going round it are shown. (*Not* to scale).

The first laboratory method of measurement of the speed of light was devised by Fizeau in 1849. The principle is illustrated in Fig. 11.3.

A light beam passes through the space between the teeth of a wheel and is reflected from a distant mirror M, and again passes through the teeth-space and enters the eye. When the wheel spins, the light gets cut off if the tooth obstructs the reflected wave. For

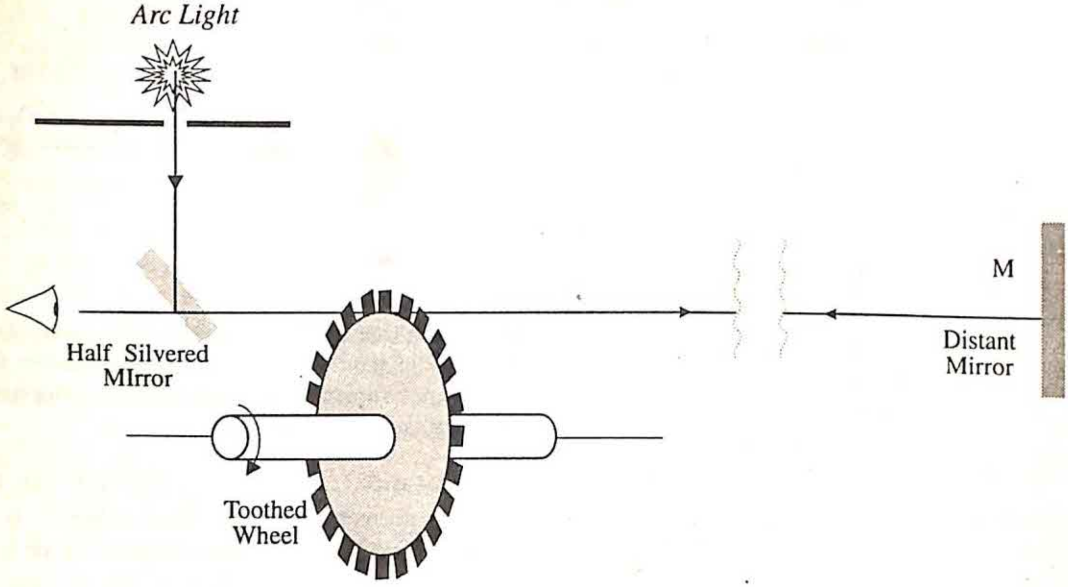


Figure 11.3: Fizeau's method for measuring the speed of light.

low angular speeds of the wheel, a tooth hasn't moved much by the time the light comes back; it goes through the gap. As the angular speed increases, the tooth could just obstruct the returning light beam. Fizeau measured the angular velocity of the wheel at which the reflected light is cut off and could then calculate the speed of light.

Very precise measurements of the speed of light were carried out by Michelson during 1926-29. An octagonal mirror M_1 is mounted on the shaft of a variable speed motor (Fig. 11.4). Light from a source S falls on the mirror M_1 at an angle of 45° and is reflected to a distant mirror M_2 which returns it. With M_1 stationary, the reflected ray strikes M_1 on the face at an angle of 45°. This light is reflected into the telescope T. When the mirror M_1 is set into rotation, light returning to it from the mirror M_2 will not, in general, be incident at an angle of 45°, and hence will not enter the telescope. When the speed of rotation of M_1 is such that the next face of the mirror, say b is in position formerly occupied by c during the time light travels from M_1 to M_2 and comes back to M_1, the light will be seen in the tele-

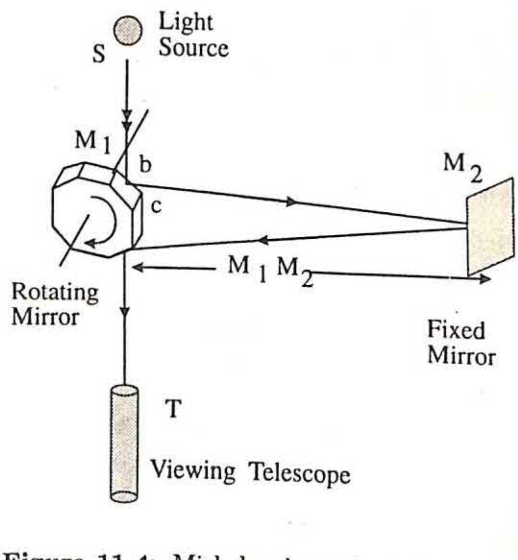


Figure 11.4: Michelson's method of measuring the speed of light.

scope. The speed of light is then

$$v = \frac{2(M_1M_2)}{t}$$

where M_1M_2 is the distance between the mirror M_1 and M_2 and t is the time of one-eighth revolution of the octagon. The speed of rotation of the spinning mirror M_1 can be measured by using a stroboscope. In Michelson's experiment, the speed of the motor was about 500 rev/s and the distance M_1 M_2 was about 35 km. (A large distance!) Michelson's measurements yielded a speed of light in air of 299729 km/s.

Modern methods of measuring the speed of light use laser beams or alternately radar signals. These methods can be modified to measure the speed of light in different materials. For example the path of light can be filled with water and the speed of light in water can be found. The speed turns out to be about three-fourth of the speed in air contradicting Newton's corpuscular theory. The speed of light in vacuum can also be measured. The speed is slightly higher in vacuum than in air.

The speed of light in vacuum or free space is taken as

$$c = 299792458 \text{ms}^{-1}$$

The speed of light in free space is considered a fundamental constant (see Section 9.4 for a discussion) and now the metre is redefined in terms of the velocity of light and the second.

11.4 Reflection of light from spherical mirrors

We know that spherical mirrors can be of two types, concave and convex. We are also familiar with the formation of images in these mirrors and their nature. Now we shall consider how the distance of the object (u), the image distance (v), and the focal length (f) are related.

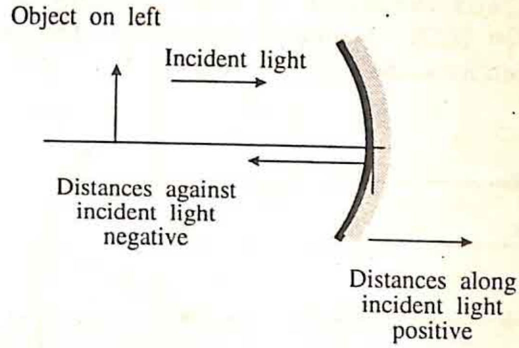


Figure 11.5: Sign convention (New cartesian)

11.4.1 Sign convention

Different sign conventions are adopted for the measurement of u, v and f. In this book we shall follow the *New Cartesian* sign convention in which all distances are measured from the pole of the mirror. The distances measured in the *same* direction as the incident light are taken as positive and those measured in the direction *opposite* to the direction of the incident light are taken as negative (Fig. 11.5).

11.4.2 Focal length and radius of curvature

Fig. 11.6 shows what happens when a parallel beam of light falls on concave and convex mirrors. The laws of reflection are applied at each point and the reflected rays are as marked. It is seen that the reflected rays converge at a point F in the case of reflection from a concave mirror. Also, the reflected rays appear to diverge from F in the case of a convex mirror. The point F is called the

principal focus of the spherical mirror. The distance f from the principal focus to the vertex or pole of the mirror is called the *focal length* of the mirror. The focal length is approximately half the radius of curvature of the mirror for small apertures i.e. $f = (1/2)\times$ (radius of curvature). This is seen from the following (Fig. 11.7).

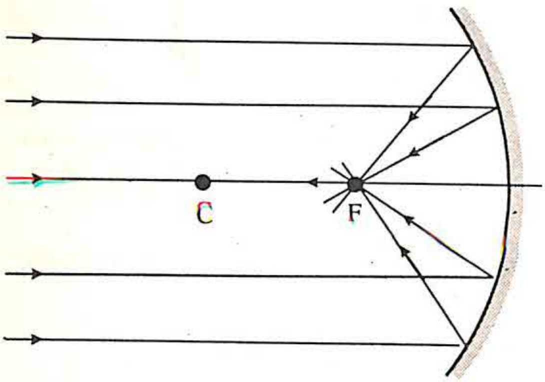


Figure 11.6(a)

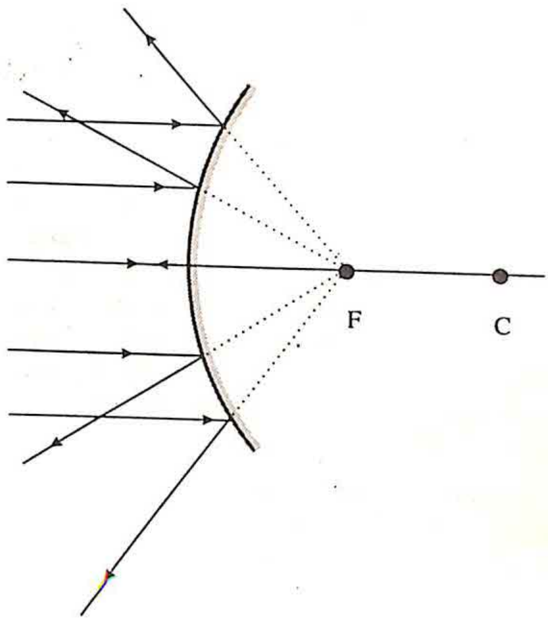


Figure 11.6(b)

Figure 11.6: Reflection of parallel light from concave and convex mirrors.

The mirror, concave or convex, has its centre of the spherical surface, marked C. The axial line or axis CV intersects the mirror at a point V called the vertex, or the pole. The distance CV is clearly the radius of curvature r of the spherical surfaces. Now consider a ray parallel to the axis incident on the mirror at the point M. Since by definition the line CM is a radius, it is perpendicular to the mirror surface at M. The angle of incidence is therefore, θ, as shown. The angle of reflection is also θ, as shown. The reflected ray hits the axis at the point F for the concave mirror. For the convex mirror, it appears to come from F. The distance FV is called the focal length f. Consider the triangle MCD where MD is the perpendicular from M on the axis. We have $\tan\theta = (MD/CD)$. From the triangle MFD, $\tan 2\theta = (MD/FD)$.

$$\text{Therefore } \frac{\tan\theta}{\tan 2\theta} = \frac{MD}{CD}\cdot\frac{FD}{MD}.$$

$$\simeq \frac{FV}{CD} = \frac{f}{CD}$$

For small θ, $\tan\theta \simeq \theta$, $\tan 2\theta \simeq 2\theta$

$$\text{Therefore } \frac{f}{CD} = \frac{1}{2} \simeq \frac{f}{\mathrm{CV}} = \frac{f}{R}$$

$$\text{or } f = \frac{1}{2}R = \frac{1}{2}\times \text{ radius of curvature.} \tag{11.4}$$

This means that the principal focus lies half way between the centre of curvature and the vertex of the mirror. We notice that the result, Eq. (11.4) for the relation between the focal length and the radius of curvature

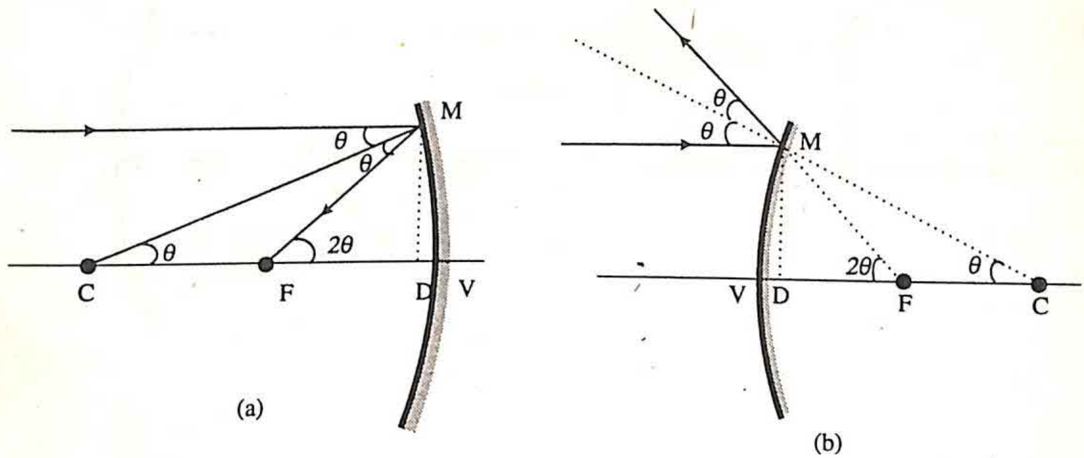


Figure 11.7: A ray parallel to the axis incident on (a) a concave spherical mirror and (b) a convex spherical mirror.

and indeed the focussing of parallel rays on to F, depends on the angle θ being small, i.e. the assumptions that the parallel rays are close to the axis, and that the distance VM is much less than the radius of curvature $R(=VC)$.

If this is not true, namely the parallel rays are not close to the axis of the spherical surface, the image is not a point, but a curve. We discuss this defect of image formation later (Section 11.15).

11.4.3 The mirror equation

A simple general relation exists for a mirror (concave or convex) between the object distance (symbol u), the image distance (symbol v) and the focal length f. This enables us to say, where the image will be formed when the object is at a known distance, and the focal length f (or radius of curvature $= 2f$) of the mirror is known. The sign convention is such that the *same* relation between u, v and f is valid whether the image is real or virtual, and whether the mirror is concave or convex! We will obtain the relation, known as the mirror equation, for a special case, namely a concave mirror producing a real image, but we shall indicate that it is true for all cases, giving other examples.

Consider a concave mirror, with the object AB as illustrated in the figure (Fig 11.8a). A ray of light from A, parallel to the axis CV of the mirror, falls on it at M. The reflected ray MA′D passed through the focal point F of the mirror. Another ray AV incident on the vertex V, is reflected and the reflected ray is VA′E. The two reflected rays meet at A′ which is therefore the *real* image of A. Similarly, by drawing rays from B, we see that its image is at B′, vertically *above* A′. The distance VB is the distance u of the object. It is negative since the direction VB is against the direction of incident light. Similarly, the distance $VB' = v$ of the image is also negative, as is the focal length $f(=VF)$. (All this is in our sign convention, called the new Cartesian sign convention). To find the relation between these, consider triangles $B'A'F$ and VMF. (We assume that V coincides with the perpendicular to the axis VC dropped from M, so that effectively MV is a straight line). The two right angled

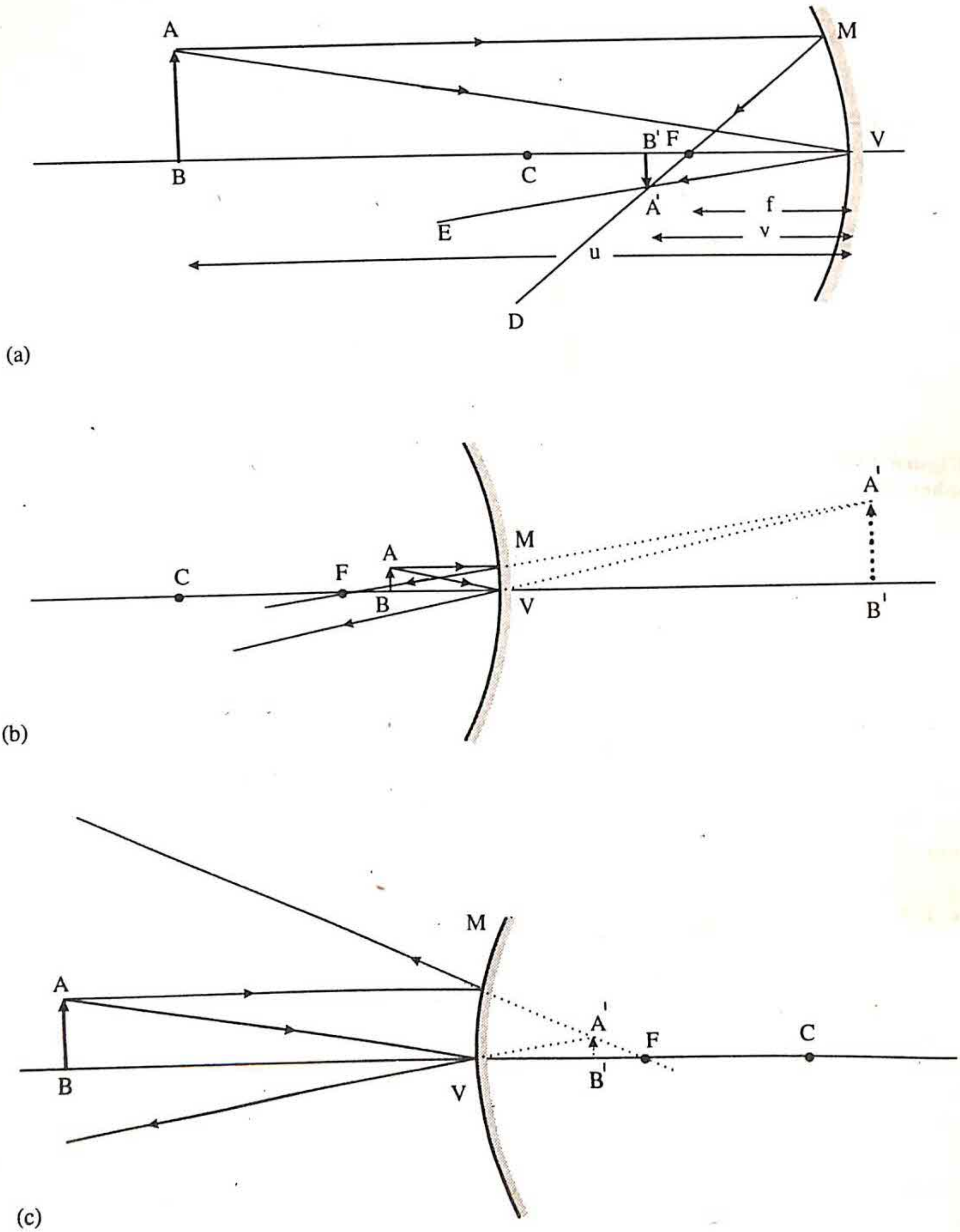


Figure 11.8: The mirror equation (a) A concave mirror with object AB farther than C. The image is reduced, inverted and real, (b) Concave mirror with object closer than F, and (c) The image formation by a convex mirror.

triangles are similar, so that

$$\frac{A'B'}{MV} = \frac{B'F}{FV} = \frac{B'V - FV}{FV} = \frac{(-v) - (-f)}{(-f)} = \frac{v - f}{f} \quad (11.5)$$

Right angled triangles $A'B'V$ and ABV are also similar, so that

$$\frac{A'B'}{AB} = \frac{B'V}{BV} = \frac{-v}{-u} \quad (11.6)$$

But from the figure, $AB = MV$, so that in Eq. (11.5),

$$\frac{A'B'}{AB} = \frac{A'B'}{MV} = \frac{v}{u}. \quad (11.7)$$

Now compare Eqs. (11.5) and (11.7) for the same quantity, namely $(A'B'/MV)$. We, therefore, have

$$\frac{v - f}{f} = \frac{v}{u}$$

or

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}. \quad (11.8)$$

This is the relation between v, u and f that we were looking for; it is called the *mirror formula*.

We notice one more new result: the ratio of the size of the image $(A'B')$ to the size of the object (AB) is given, according to the Eq. (11.6), by (v/u). This ratio is, naturally, called magnification. So we have the result

$$\text{magnification} = \frac{\text{size of image}}{\text{size of object}} = \frac{v}{u} \quad (11.9)$$

In the case shown, the image is smaller than the object, i.e magnification is less than one; the image inverted, and real as well. A concave mirror forms real, inverted, reduced images when the object is a distance longer than the radius.

Images can be located by tracing a few rays from the object, that parallel to the axis (they are reflected through F), the one passing through F (they are reflected parallel to the axis), the one that is incident at the pole (reflected with angle of reflection equal to the angle of incidence) and the one that passes through the centre C (reflected back in the same direction).

We shall not prove in detail the mirror equation Eq. (11.8) for all possible cases, but indicate through figures the two other (different) situations. In Fig (11.8b), we show what happens for a concave mirror when the object is at a distance VB less than the focal length. The image is virtual, so that $VB'(= v)$ is positive.

Again, considering the similar triangles MVF and $A'B'F$, we have

$$\frac{A'B'}{MV} = \frac{FB'}{FV} = \frac{FV + VB'}{FV} = \frac{-f + v}{-f} \quad (11.10)$$

where the last equation is in terms of v (positive and equal to VB') and f (negative and equal to FV in magnitude). Again, considering the similar triangles ABV and $A'VB'$, we have

$$\frac{A'B'}{MV} = \frac{A'B'}{MV} = \frac{v}{-u} \quad (11.11)$$

Comparing Eqs. (11.10) and (11.11), we have *same* mirror equation, Eq. (11.8). The result for the case of a virtual image formed by a convex mirror (Fig 11.8c) is proved in a manner very similar to that above. In this case, the object distance u is negative. But the image distance v and the focal length are both positive, because they are both behind the mirror, and in the direction of incident light from V. The image is always erect and virtual. Also, the image size is always less than that of the object. The mirror relation Eq. (11.8) follows from the same kind of arguments as above.

Example 11.1: An object is placed (i) 10

cm, (ii) 5 cm in front of a concave mirror of radius of curvature 15 cm. Calculate the position, nature and magnification of the image in each case.

Answer: The focal length $f = -15/2 = -7.5$cm.

(i) The object distance $u = -10$ cm.

$$\text{Since } \frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

$$\frac{1}{v} + \frac{1}{-10} = \frac{1}{-7.5}$$

$$\text{or } v = \frac{10 \times 7.5}{7.5 - 10} = \frac{10 \times 7.5}{-2.5} = -30\text{cm}$$

The image is 30 cm from the mirror on the object side.

$$\text{Also magnification } M = \frac{v}{u} = \frac{-30}{-10} = 3.$$

(ii) The object distance $u = -5$cm

Since

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

$$\frac{1}{v} + \frac{1}{-5} = \frac{1}{-7.5}$$

or

$$v = \frac{5 \times 7.5}{7.5 - 5} = 15\text{cm}.$$

This image is 15cm at the back of the mirror (i.e., not on the object side). It is a virtual image.

$$\text{Magnification } M = \frac{v}{u} = \frac{15}{-5} = -3.$$

The minus sign indicates that the magnified image is the same way up as the object.

11.5 Refraction

11.5.1 Law of refraction

We have already discussed the passage of a plane wavefront of light from one medium to the second, in Section (10.2.2), Chapter 10. The ray direction is perpendicular to the plane of the wavefront. We found there that the angle of incidence i of the incident ray (angle of the ray with respect to the normal) and the angle of refraction r of the refracted ray are related by Snell's law, namely

$$\frac{\sin i}{\sin r} = n \qquad (11.12)$$

where n is the refractive index of the second medium with respect to the first. Interestingly, the wave theory identifies the refractive index with the ratio of velocities of light v_1 and v_2 in the two media, namely,

$$n = (v_1/v_2) \qquad (11.13)$$

(The relation Eq. (11.12) was discovered experimentally centuries ago, among others by Snell, a Dutch scientist. The connection between refractive index and the velocity of light is the work of Huygens' (see Chapter 10).

Equation (11.12) can be formulated more precisely as follows: The angles of incidence and refraction are related by the following law:

$$\frac{\sin i}{\sin r} = n_{21} = \frac{n_2}{n_1} = \frac{v_1}{v_2} \qquad (11.14)$$

where n_{21} is the refractive index of medium 2 (into which the light ray is refracted) with respect to medium one (from which the light is incident). Because of the result, Eq. (11.13), n_{21} can be always written as the ratio of absolute refractive indices n_2 and n_1 as in Eq. (11.14). For example n_2 is just the ratio of the velocity of light in vacuum to that in the medium 2. Often we use the phrase 'refractive index of a medium. We clearly mean the

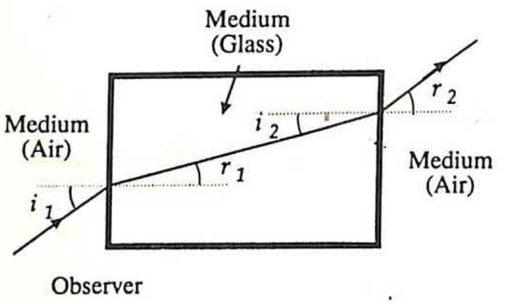


Figure 11.9: Lateral shift of a ray refracted through a parallel sided slab.

refractive index of one medium with respect to the other.

We discuss some examples and consequences of refraction in this section.

11.5.2 Some examples of refraction

(i) Lateral shift

When a ray of light travels from one medium to a second medium and then to a third medium, refraction occurs at both the interfaces. Fig. 11.9 traces the path of light as it travels through a parallel glass slab.

Suppose the medium on both sides of the glass slab is the same. In what direction will the ray of light emerge from the glass slab? We can easily answer this question.

$$\frac{\sin i_1}{\sin r_1} = n$$

$$\frac{\sin i_2}{\sin r_2} = \frac{1}{n}$$

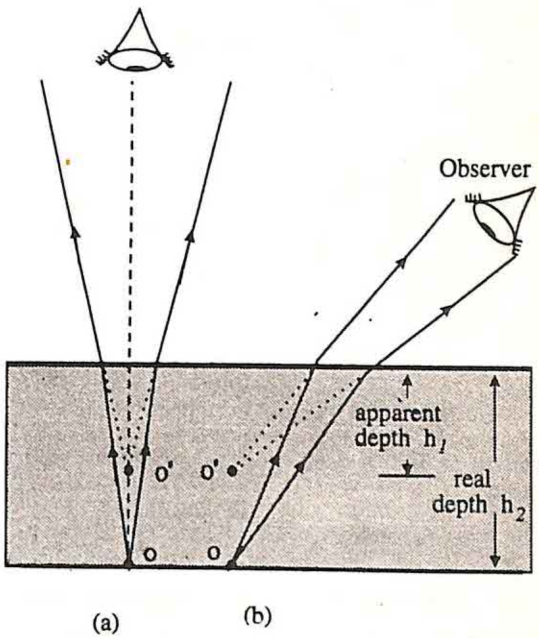
But $i_2 = r_1$.

Therefore

$$\frac{\sin i_1}{\sin r_1} = \frac{\sin r_2}{\sin r_1}$$

or $i_1 = r_2$.

The emergent ray is parallel to the incident ray, but is displaced.



(a) for normal incidence
(b) for oblique incidence

Figure 11.10: Apparent depth for (a) normal and (b) oblique incidence

(ii) Apparent depth

The apparent depth of objects is lesser in denser media than the actual depth (Fig. 11.10).

The tracing of the refracted beam makes us visualize the object O as apparently being at O′. In fact, as you can work out that

$$\frac{\text{actual depth}}{\text{apparent depth}} = \text{refractive index of the medium.}$$

(iii) Delayed sunset and advanced sunrise

The sun is visible before actual sunrise and after actual sunset because of atmospheric refraction (Fig. 11.11). By actual sunrise we

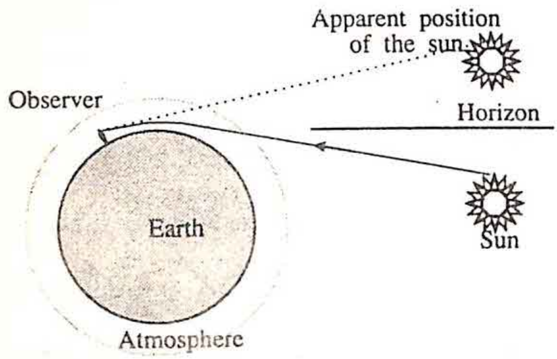


Figure 11.11: Refraction effect at sunrise and sunset.

mean the *sunrise* actual crossing the horizon by the sun.

We show the actual and apparent positions of the sun with respect to the horizon. The figure is highly exaggerated to show the effect. Refractive index of air with respect to free or outer space is 1.00029. Due to this the apparent shift in the direction of the sun is by about $(1/2)°$ and the corresponding time difference between actual sunset and apparent sunset is about 2 min. The apparent flattening of the sun at sunset and sunrise also is due to the same reason.

11.5.3 Total internal reflection

(i) Total internal reflection and the critical angle

When a ray of light travels from a denser medium to a rarer medium (i.e. one with a smaller refractive index), the refracted ray is bent away from the normal, e.g. AO_1 B in Fig. 11.12. The incident AO_1 is partially reflected (O_1C) and partially transmitted (O_1B) or refracted, the angle of refraction being larger than the angle of incidence. As the angle of incidence increases, so does the angle of refraction, till for the

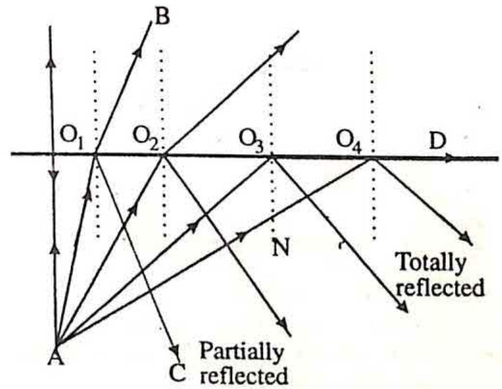


Figure 11.12: Total internal reflection.

ray AO_3, the angle of refraction is 90°. The refracted ray is bent so much away from the normal that it grazes the surface or interface between the two media. This is the ray AO_3D. If the angle of incidence is increased still further (e.g. the ray AO_4) refraction is not possible, and the incident ray is totally reflected. This is called *total* internal reflection.

The critical angle of incidence, $\angle AO_3N$ is such that the angle of refraction is 90°. We see from Snell's law (Eq. 11.12) that if the relative refractive index is less than one, then since the maximum value of sin r is unity, there is an upper limit to the value of sin i for which the law can be satisfied. This is $i = i_c$ such that

$$\sin i_c = n$$

For larger values of i, i.e. larger values of $\sin i$, Snell's law of refraction cannot be satisfied for any value of r (since the maximum value of sin r is unity). There is no refraction.

The critical angle for total internal reflection when light is incident on a rarer medium 2 from a denser medium 1 is given from Snell's law by

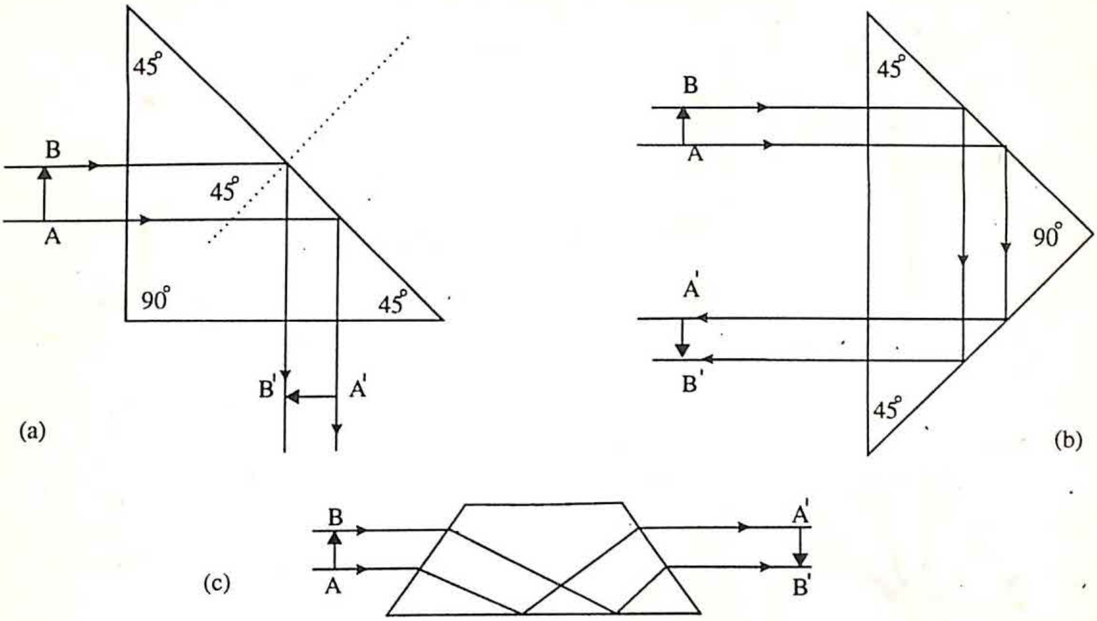


Figure 11.13: Prisms using total internal reflection to bend rays by 90° and 180°.

Table 11.2: Critical angles for some transparent media.

Substance	Refractive index	Critical angle
Water	1.33	48.75°
Crown glass	1.52	41.14°
Dense flint glass	1.65	37.31°
Diamond	2.42	24.41°

$$\frac{\sin i_c}{\sin(\pi/2)} = \sin i_c = n_{21} = \frac{n_2}{n_1} \qquad (11.15)$$

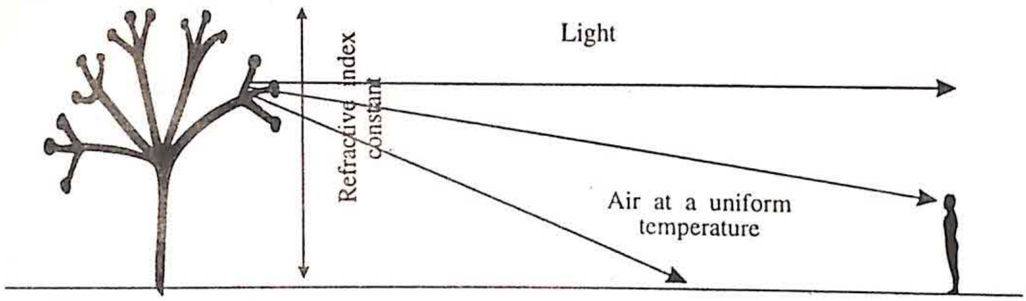
since by definition the angle of refraction for critical incidence is $(\pi/2)$ radians or 90°. Some typical critical angles are listed in Table 11.2.

Total internal reflection is the main cause of the brilliance of diamonds. Its critical angle is very small, so that once light gets into diamond, it is very likely to be totally reflected internally. By cutting the diamond suitably, multiple internal reflections can be made to occur.

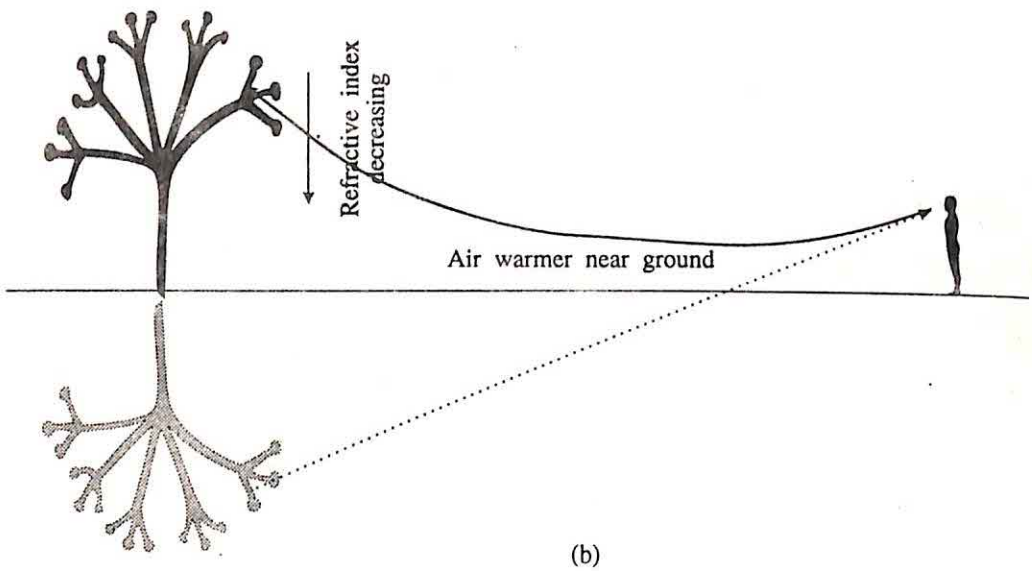
Prisms make use of total interval reflections to bend light by 90° (Fig. 11.13(a)) or by 180° (Fig. 11.13(b)), or to invert images without changing their size (Fig. 11.13(c)). In the first two cases, the critical angle i_c for the material of the prism must be less than 45°. We see from Table 11.2 that this is true for both crown glass and dense flint glass.

(ii) Mirage

On still summer days, the air near the ground may become hotter than air further up. Now the refractive index of air increases with its density. Hotter air is less dense, and so has a smaller refractive index than cooler air. So, light from a tall object such as a tree passes through a medium whose refractive index decreases towards the ground. Thus a ray of light from such an object gets bent and is totally internally reflected. This is shown in Fig. 11.14b. This light reaches the observer in the direction shown,. He naturally assumes that it is reflected from the ground, say by a pool of water there. Such inverted images of distant high objects cause



(a)



(b)

Figure 11.14: (a) A tree as seen by an observer when the air above the ground is at a uniform temperature. (b) When the air close to the ground is hot, light bends gradually as shown, undergoing total internal reflection, and the apparent image of the tree may mislead the observer into thinking that there is a pool of water in front of the tree!

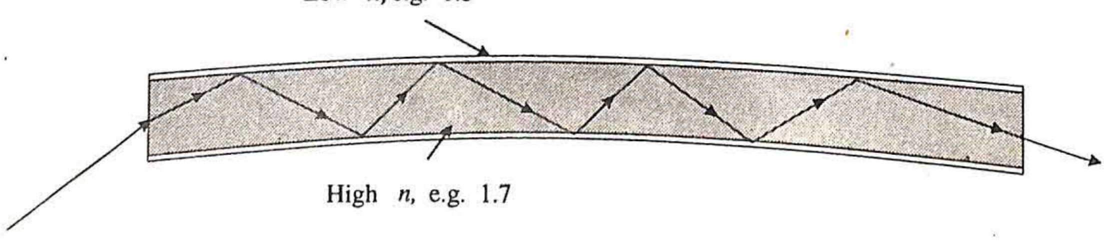


Figure 11.15: An optical fibre

the optical illusion called a *mirage*, specially common in hot deserts.

(iii) Optical fibres

Optical fibres too make use of the same total internal reflection principle. Optical fibres consist of many long high quality glass/quartz fibres. Each is coated with a thin layer of a material of lower refractive index. For example, a strand can be 0.0001 cm in diameter, with the refractive index of the main fibre being 1.7 and that of the coating 1.5.

When the light is incident on one end of the fibre at a small angle, the light passes inside, undergoes repeated total internal reflections along the fibre and finally comes out (Fig. 11.15). The angle of incidence is always larger than the critical angle of the fibre material with respect to its coating. Even if the fibre is bent, the light can easily travel through along the fibre.

A bundle of optical fibres can be put to several uses. It can be used as a 'light pipe' in medical and optical examination. It can also be used for optical signal transmission. Optical fibres have also been used for transmitting and receiving electrical signals which are converted to light by suitable transducers. The main requirement is that there be very little absorption of light as it travels for long distances inside the optical fibre. This has been achieved by purification and special preparation of materials such as quartz. In silica glass fibres, it is possible to transmit more than 95% of the light over a fibre length of 1 km. (Compare with what you expect for a block of ordinary window glass 1 km thick).

11.5.4 Refraction in a prism

Let us consider the passage of a light ray through a glass prism (Fig. 11.16). A ray

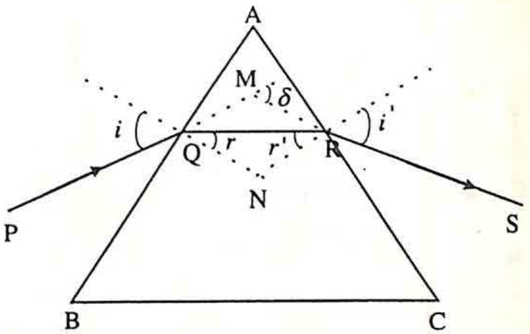


Figure 11.16: Light ray passing through a triangular glass prism.

PQ is incident on the face AB at an angle i, and is refracted along QR inside the prism. The angle of refraction is r. The ray QR finally emerges along RS, away from the normal to the face AC. At this face, angle of incidence is r' and the angle of refraction or emergence is i'. The deviation of the emergent ray RS from the incident direction PQ is δ (see figure). We now show that under certain conditions, the refractive index of the material of the prism can be related to the angle of deviation δ, and the angle A of the prism.

From the quadrilateral AQNR, which has right angles at the vertices Q and R,

$$A + Q\hat{N}R = 180°.$$

From the triangle QNR,

$$r + r' + Q\hat{N}R = 180°$$

Comparing these two equations, we find that

$$A = r + r' \tag{11.16}$$

From the triangle MQR,

$$\begin{aligned}\delta &= \text{deviation at the first surface} \\ &\quad + \text{deviation at the second surface} \\ &= (i - r) + (i' - r') \\ &= i + i' - A \end{aligned} \tag{11.17}$$

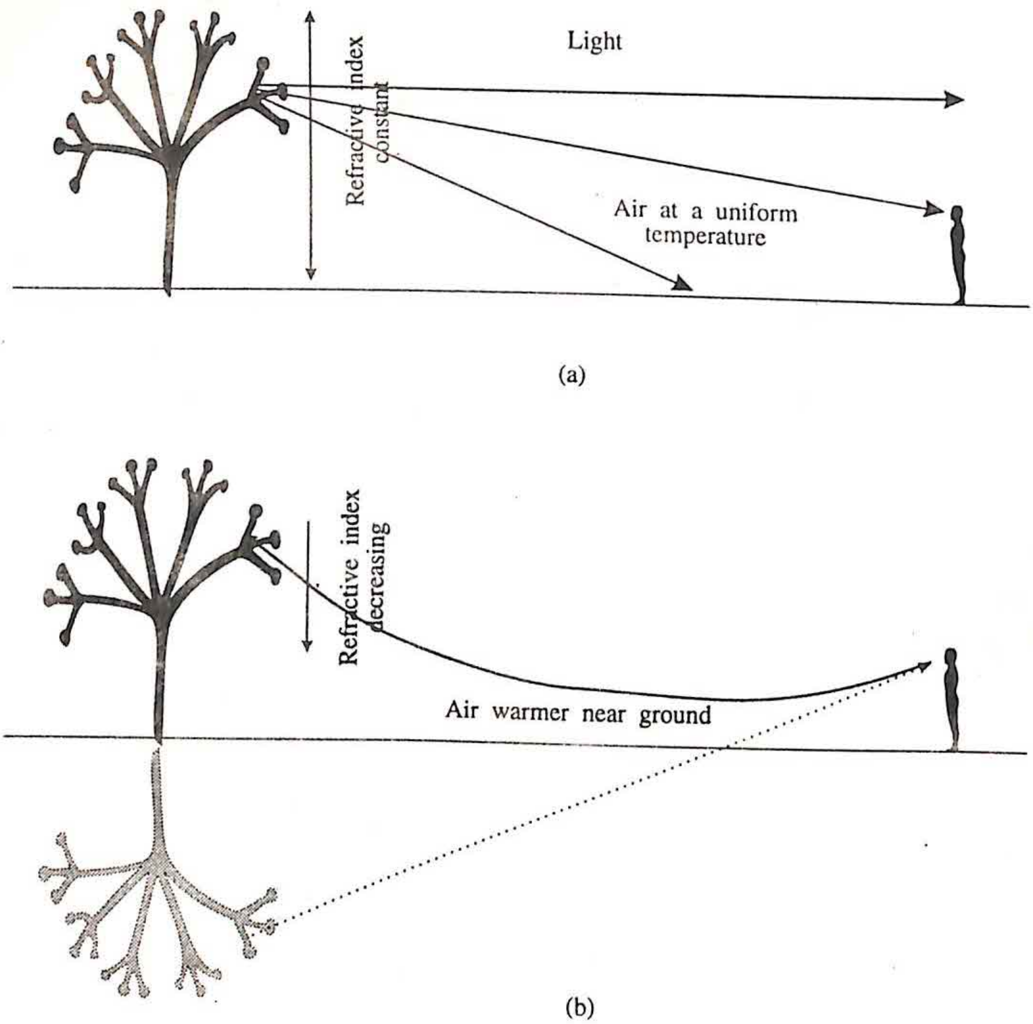


Figure 11.14: (a) A tree as seen by an observer when the air above the ground is at a uniform temperature. (b) When the air close to the ground is hot, light bends gradually as shown, undergoing total internal reflection, and the apparent image of the tree may mislead the observer into thinking that there is a pool of water in front of the tree!

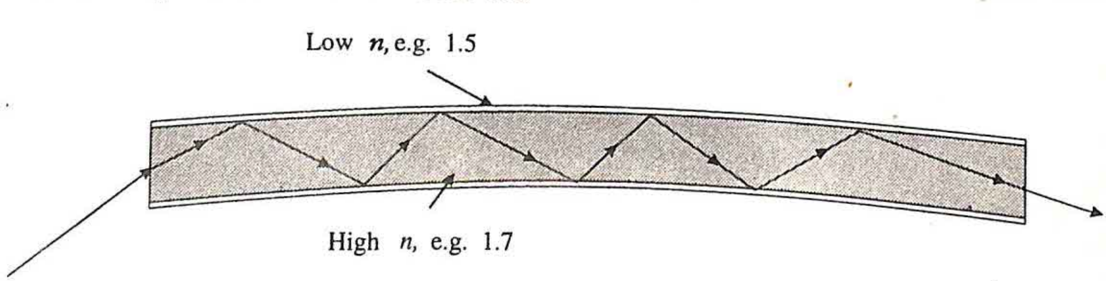


Figure 11.15: An optical fibre

the optical illusion called a *mirage*, specially common in hot deserts.

(iii) Optical fibres

Optical fibres too make use of the same total internal reflection principle. Optical fibres consist of many long high quality glass/quartz fibres. Each is coated with a thin layer of a material of lower refractive index. For example, a strand can be 0.0001 cm in diameter, with the refractive index of the main fibre being 1.7 and that of the coating 1.5.

When the light is incident on one end of the fibre at a small angle, the light passes inside, undergoes repeated total internal reflections along the fibre and finally comes out (Fig. 11.15). The angle of incidence is always larger than the critical angle of the fibre material with respect to its coating. Even if the fibre is bent, the light can easily travel through along the fibre.

A bundle of optical fibres can be put to several uses. It can be used as a 'light pipe' in medical and optical examination. It can also be used for optical signal transmission. Optical fibres have also been used for transmitting and receiving electrical signals which are converted to light by suitable transducers. The main requirement is that there be very little absorption of light as it travels for long distances inside the optical fibre. This has been achieved by purification and special preparation of materials such as quartz. In silica glass fibres, it is possible to transmit more than 95% of the light over a fibre length of 1 km. (Compare with what you expect for a block of ordinary window glass 1 km thick).

11.5.4 Refraction in a prism

Let us consider the passage of a light ray through a glass prism (Fig. 11.16). A ray

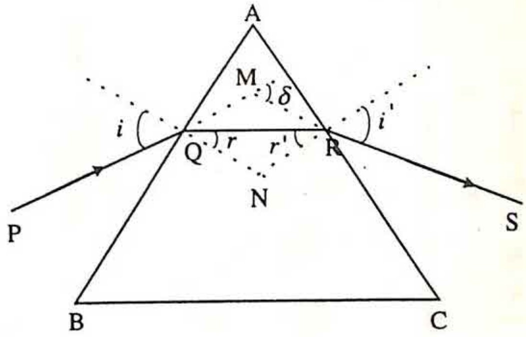


Figure 11.16: Light ray passing through a triangular glass prism.

PQ is incident on the face AB at an angle i, and is refracted along QR inside the prism. The angle of refraction is r. The ray QR finally emerges along RS, away from the normal to the face AC. At this face, angle of incidence is r' and the angle of refraction or emergence is i'. The deviation of the emergent ray RS from the incident direction PQ is δ (see figure). We now show that under certain conditions, the refractive index of the material of the prism can be related to the angle of deviation δ, and the angle A of the prism.

From the quadrilateral AQNR, which has right angles at the vertices Q and R,

$$A + Q\hat{N}R = 180°.$$

From the triangle QNR,

$$r + r' + Q\hat{N}R = 180°$$

Comparing these two equations, we find that

$$A = r + r' \tag{11.16}$$

From the triangle MQR,

$$\begin{aligned}\delta &= \text{deviation at the first surface} \\ &\quad + \text{deviation at the second surface} \\ &= (i - r) + (i' - r') \\ &= i + i' - A \end{aligned} \tag{11.17}$$

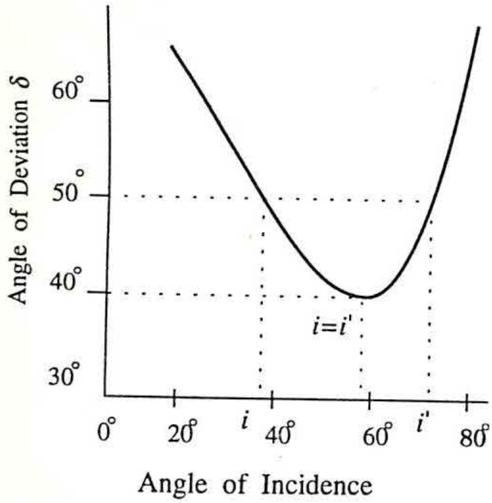


Figure 11.17: Plot of angle of deviation δ versus angle of incidence i for a triangular prism.

As the angle of incidence is increased, the angle of deviation first decreases, reaches a minimum and then increases (Fig. 11.17). For any angle of deviation, the angle of incidence has two values i and i'. At minimum deviation $\delta = D$, $i = i'$ or the incident ray and the emergent rays are symmetrical with respect to the refracting faces and the refracted ray in the prism is parallel to the base.

Since $i = i'$, one has $r = r'$ so that from Eq. (11.16), we have

$$A = 2r : \ r = A/2. \tag{11.18}$$

Similarly, from Eq. (11.17),

$$D = 2i - A \text{ or } i = \frac{(A+D)}{2} \tag{11.19}$$

From Snell's law,

$$n_1 \sin\frac{(A+D)}{2} = n_2 \sin\frac{A}{2}$$

or

$$\frac{n_1}{n_2} = \frac{\sin[(A+D)/2]}{\sin(A/2)} \tag{11.20}$$

This equation provides an accurate method for measuring the refractive index of a transparent prism in terms of two angles which can be measured precisely.

11.6 Refraction at spherical surfaces and by lenses

11.6.1 Refraction at a spherical surface

We now consider refraction of light not at a plane surface, but at a spherical surface (Fig. 11.18). This is the first step towards a description of lenses, which are transparent optical materials bounded by *two* spherical surfaces. As you know, lenses, convex and concave, are very widely used singly and in combination in a large number of optical instruments.

Consider a spherical surface NM (Fig. 11.18) which is the interface between two media, one of refractive index n_1 (to the left) and the other of refractive index n_2. Its centre of curvature is at C. The radius of curvature $MC = |R|$. There is an object O in the medium 1. A ray of light from O is incident on the interface at the point N; the angle of incidence is i. This ray is refracted and intersects the axis at I. This is where the image is formed. The axis OMC is the line from O perpendicular to the spherical surface, and extended in both directions. A ray along OM passes through without deviation (normal incidence). Two rays from O, namely OMCI and ONI intersect at I which is therefore the image. We now relate the object distance u, and the image distance v, with the properties of the surface and the medium, using Snell's law. The relation is easily obtained for small apertures, or small ratios (MN/OM). (i.e. for (MN/OM) << 1).

The angle of incidence i is given from Δ NOC as its exterior angle, i.e. one has

$$i = N\hat{O}M + N\hat{C}M$$

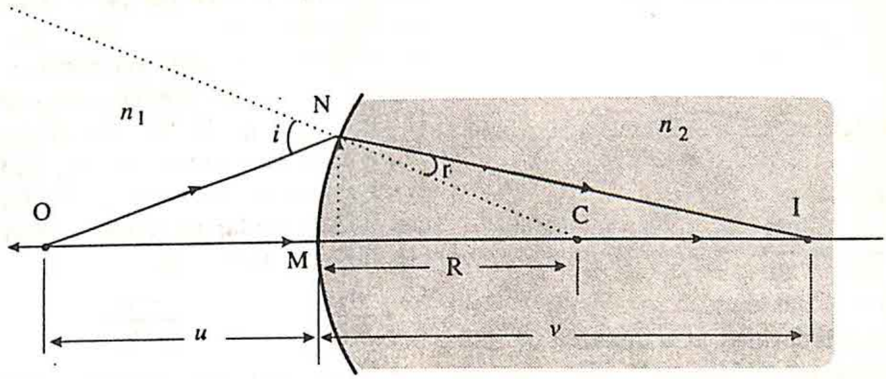


Figure 11.18: Refraction at a spherical surface separating two media.

$$\tan N\hat{O}M = \frac{MN}{OM} = N\hat{O}M$$
(for small angles)

$$\tan N\hat{C}M = \frac{MN}{MC} = N\hat{C}M$$
(for small angles)

$$\text{Hence} \quad i = \frac{MN}{OM} + \frac{MN}{MC} \tag{11.21}$$

Similarly, $r = N\hat{C}M - N\hat{I}M$

$$\tan N\hat{I}M = \frac{MN}{MI} = N\hat{I}M$$
(for small angles)

$$\text{Therefore} \quad r = \frac{MN}{MC} - \frac{MN}{MI} \tag{11.22}$$

Now i and r are related by the refractive index relation

$$n_1 \sin i = n_2 \sin r.$$

Hence for small angles, using Eq. (11.21) and (11.22)

$$n_1\left(\frac{MN}{OM} + \frac{MN}{MC}\right) = n_2\left(\frac{MN}{MC} - \frac{MN}{MI}\right)$$

or

$$\frac{n_1}{OM} + \frac{n_2}{MI} = \frac{(n_2 - n_1)}{MC} \tag{11.23}$$

Here OM, MI and MC are lengths, namely positive numbers.

As in the case of mirrors, we shall follow the *New Cartesian* sign convention for u, v and f. All distances are measured from the centre of the refracting surface on the optic axis. The distances measured in the same direction as the incident light are taken as positive. The distances measured in the direction opposite to the direction of the incident light are taken as negative. With this convention, u is negative, and v and R are positive. We then have $u = -OM, v = MI, R = MC$. The above equation becomes

$$\frac{n_1}{-u} + \frac{n_2}{v} = \frac{n_2 - n_1}{+R}$$

or

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R} \tag{11.24}$$

This formula is to be used for refraction through spherical interfaces.

Example 11.2: Light from a point source in air falls on a spherical glass surface ($n = 1.5$, radius of curvature = 20 cm). The distance of the light source from the glass surface is 100 cm. At what position is the image formed?

Answer: We use the formula

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R}.$$

Here

$u = -100\text{cm}, v = ?, R = +20\text{cm}, n_1 = 1$, and $n_2 = 1.5$.

We then have

$$\frac{1.5}{v} + \frac{1}{100} = \frac{0.5}{20}$$

or $\quad v = +100$ cm.

The image is formed at a distance of 100 cm from the glass surface, in the direction of incident light.

11.6.2 Thin lens formula

We now consider how image formation occurs in thin concave and convex lenses. We will find a simple relationship between the object distance (u), the image distance (v) and the focal length (f) of the lens (Fig. 11.19a). In Fig. 11.19, the formation of the image is shown in two steps. First the spherical interface corresponding to the side ABC (centre of curvature C_1) of the lens forms the image at I_1 (Fig. 11.19b). This image at I_1 acts as an object for the spherical interface ADC (centre of curvature at C_2). The second image is formed at I_2 (Fig. 11.19c). I_2 is the effective position of the image corresponding to the lens ABCDA as a whole.

We can use the spherical interface equation obtained above, Eq. (11.24).

For the interface ABC (Fig. 11.19b),

$$\frac{n_1}{OB} + \frac{n_2}{BI_1} = \frac{(n_2 - n_1)}{BC_1} \qquad (11.25)$$

Here we have considered the ray $ON_1N_2I_1$ which is refracted at N_1 from a medium of refractive index n_1 into a medium of refractive index n_2. Equations similar to Eq. (11.21) and (11.22) are obtained for i_1 and r_1 by drawing the appropriate triangles (Fig.(11.19b)). The ray N_1N_2 travelling in the medium of refractive index n_2, refracts at N_2 (i.e. the spherical surface ADC in Fig.(11.19a)) to form a real image at I_2 (Fig. 11.19c). In Fig. 11.19c, the angles of incidence and refraction i_2 and r_2, and the relevant triangles are shown. We find, using arguments similar to those for deriving Eq. (11.21) and (11.22), that

$$-\frac{n_2}{DI_1} + \frac{n_1}{DI_2} = \frac{n_2 - n_1}{DC_2} \qquad (11.26)$$

Adding these two equations (Eqs. (11.25) and (11.26)), and assuming $BI_1 = DI_1$,

$$\frac{n_1}{OB} + \frac{n_1}{DI_2} = (n_2 - n_1)\left(\frac{1}{BC_1} + \frac{1}{DC_2}\right) \qquad (11.27)$$

If O was at infinity, the image would have been at the principal focus ($OB = \infty, DI_2 = f$).

$$\frac{n_1}{f} = (n_2 - n_1)\left(\frac{1}{BC_1} + \frac{1}{DC_2}\right). \qquad (11.28)$$

Making this substitution in the previous equation,

$$\frac{n_1}{OB} + \frac{n_1}{DI_2} = \frac{n_1}{f}$$

or

$$\frac{1}{OB} + \frac{1}{DI_1} = \frac{1}{f}. \qquad (11.29)$$

So far, we have expressed our results (Eq. (11.24) to (11.28)) in terms of distances such as OB or DI_2 which are always positive.

Now we incorporate the sign convention. With the *new cartesian sign convention* u is negative and is equal to $-OB$, and $v = +DI_2$. Hence the Eq. (11.29) becomes

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f} \qquad (11.30)$$

Also $BC = +R_1, DC_2 = -R_2$ (neglecting the thickness of the lens.) R_2 is negative because it is a length from the pole of the

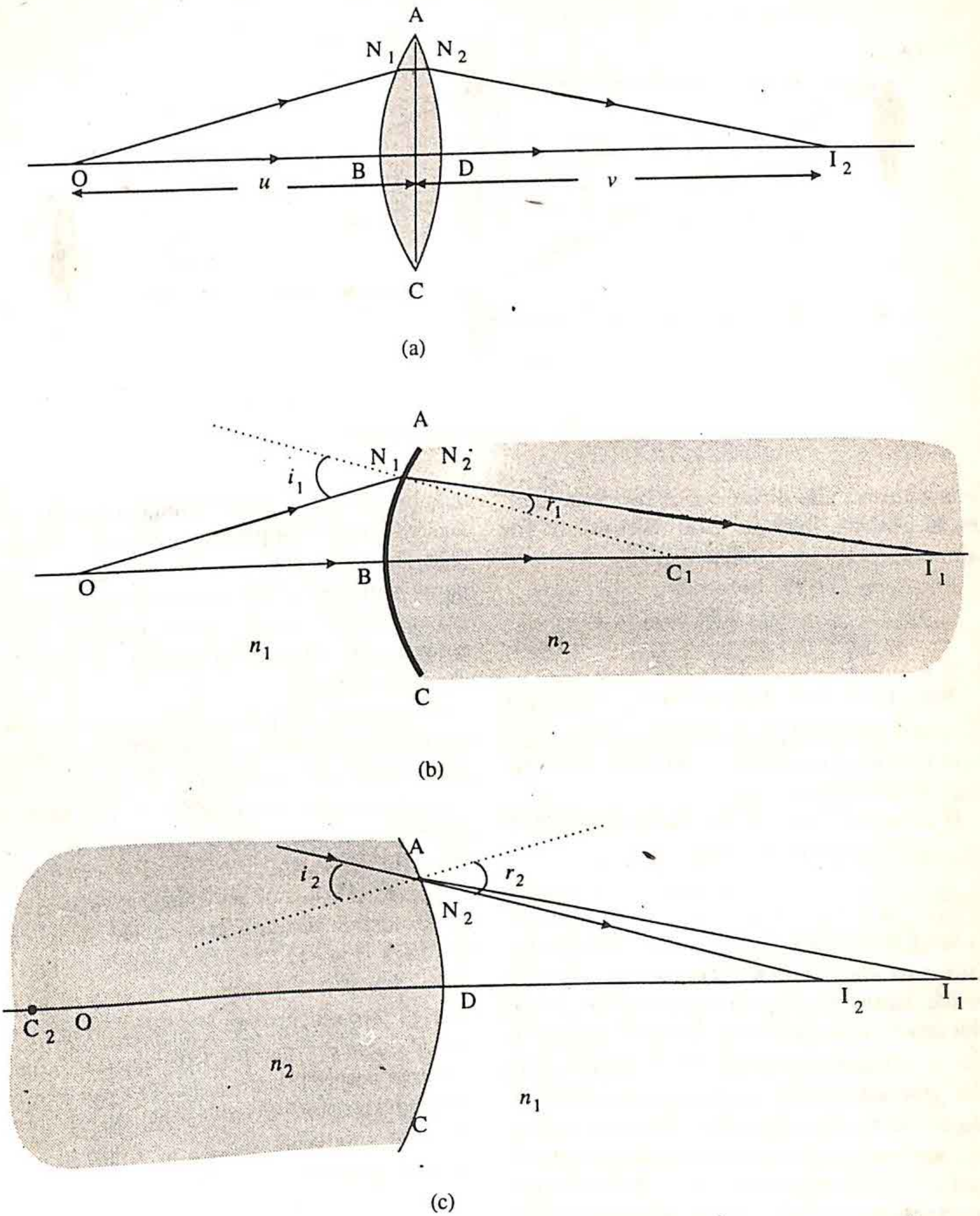


Figure 11.19: Thin lens formula. (a) The object, the double convex lens and the image (b) Refraction at the first spherical surface. (c) Refraction at the second spherical surface.

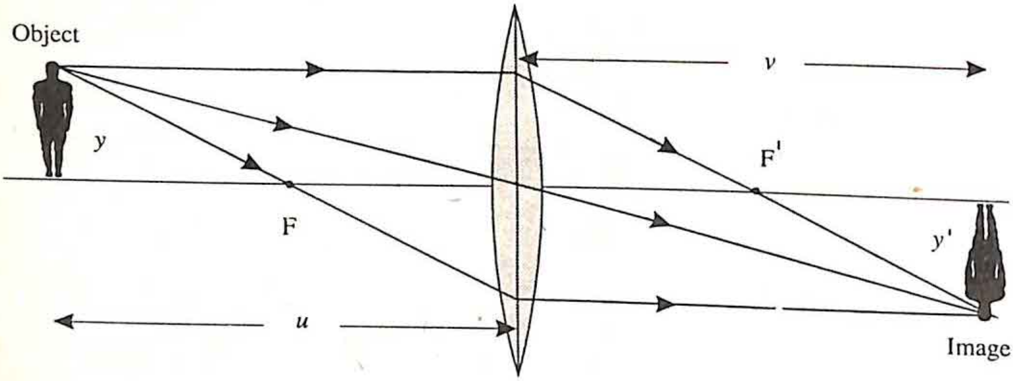


Figure 11.20: Tracing rays through a lens.

lens *against* the direction of incident light. R_1 is positive because it is distance in the same direction as incident light.

Equation (11.28) becomes

$$\frac{n_1}{+f} = (n_2 - n_1)\left[\frac{1}{+R_1} - \frac{1}{R_2}\right] \qquad (11.31)$$

Note that f is positive for a converging lens and negative for a diverging lens. This lens formula is applicable for both converging and diverging leńss.

It is easy to show, as for the spherical mirror, that the magnification

$$M = \frac{v}{u}. \qquad (11.32)$$

We have obtained above the very important and widely used formula, Eq. (11.30) which connects the object distance u and the image distance v to the focal length f. Eq. (11.31) connects the focal length f to the properties of the lens and the surrounding medium, namely their refractive indices n_2, and n_1, and the radii of curvature R_1 and R_2. The result has been obtained for a special case, namely convex (double convex) lens producing a real image. By proceeding in a way similar to that done above, we can show for any other case as well (e.g. convex lens producing a virtual image, or a concave lens which always produces a virtual image) that the same formula follows, *if the new cartesian sign convention is used.*

Lenses are called double convex, plano convex, etc., depending on the curvature of the two surfaces.

A lens has *two* principal foci, one for each direction of incident light. They are often given separate symbols, e.g. F_1 and F_2. The corresponding focal lengths are f_1 and f_2; using the definition Eq. (11.31), we find that they are equal.

The formation of an image by a lens can be discussed through tracing the path of a few rays (Fig. 11.20). For example, the ray from the object parallel to the axis is refracted to pass through the focal point or focus F'. The ray passing through the focus F emerges parallel to the axis. The ray passing through the pole of the lens (the point where the lens intersects the optic axis) emerges without change in direction.

11.6.3 Power of a lens

The focal length of a lens greatly affects the nature and size of the image of a given ob-

ject. The power (P) of a lens is a simple characterization of the focal properties of a lens.

$$P = \frac{1}{f} \tag{11.33}$$

The unit of power of a lens is m^{-1}, also called as dioptre. 1 dioptre is the power of a lens whose principal focal length is 1 m. The power of a converging lens is positive and that of a diverging lens is negative.

From the lens formula

$$\frac{n_1}{f} = (n_2 - n_1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

the power

$$P = \left(\frac{n_2}{n_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{11.34}$$

where R_1 and R_2 are the radii of curvature (with appropriate signs according to the sign convertion) of the spherical surfaces of the lens and n_2/n_1 is the refractive index of lens medium with respect to outside medium.

Example 11.3: (i) If $f = +0.5$ m, what is the power of the lens?

(ii) The radii of curvature of the faces of a double convex lens are 10 cm and 15 cm. Its focal length is 12 cm. What is the refractive index of glass?

(iii) A convex lens has 20 cm focal length in air. What is the focal length in water? (Refractive index of air-water = 1.33, refractive index for air-glass is 1.5).

Answer: (i) Power = +2 dioptre.

(ii) Here we have $f = +12\text{cm}, R_1 = +10\text{cm}, R_2 = -15$ cm.

Refractive index of air medium is taken as unity.

The lens formula is

$$\frac{1}{f} = \left(\frac{n_2}{n_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

The sign convention has to be incorporated for f, R_1 and R_2.

Substituting the values, we have

$$\frac{1}{12} = (n-1)\left(\frac{1}{10} - \frac{1}{-15}\right)$$

from which $n = 1.5$.

(iii) We have the lens formula

$$\frac{1}{f} = \left(\frac{n_2}{n_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

For a glass lens in air, $n_2 = 1.5, n_1 = 1, f = +20$ cm.

$$\frac{1}{20} = 0.5\left[\frac{1}{R_1} - \frac{1}{R_2}\right]$$

For the same glass lens in water, $n_2 = 1.5, n_1 = 1.33$. Therefore

$$\frac{1.33}{f} = (1.5 - 1.33)\left[\frac{1}{R_1} - \frac{1}{R_2}\right].$$

Combining these two equations, we find $f = +78.2$ cm.

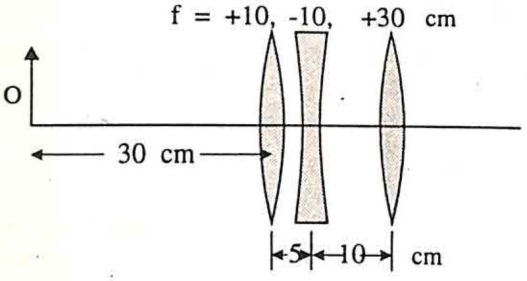
11.6.4 Lens combinations

Two or more lenses can be used in combination. This can be either for magnifying an image, making it erect or for any other purpose. The location, size and nature of the final image can be found by using either the lens equation or the ray diagram. In both the methods, we first determine the image formed by the first lens and then consider this image as the object for the second lens. If there are more than two lenses used in combination, each magnifies the image formed by the preceding lens. Hence the total magnification (M) is a product of the magnifications of the individual lenses, i. e.,

$$M = M_1 \times M_2 \times M_3 ... \tag{11.35}$$

(If the size of the image is less than the object size, then M is less than unity). We illustrate the working of such a combination through an example.

Example 11.4: Find the position of the image formed by the lens combination given in the Fig. (a) below:



Answer: Image formed by the first lens

$$\frac{1}{v_1} - \frac{1}{u_1} = \frac{1}{f_1}$$

$$\frac{1}{v_1} - \frac{1}{-30} = \frac{1}{10}$$

or $v_1 = 15\text{cm}$

The image formed by the first lens serves as the object for the second. This is at a distance of $15 - 5 = 10$ cm to the right of the second lens. It is a virtual object.

$$\frac{1}{v_2} - \frac{1}{10} = \frac{1}{-10}$$

or $v_2 = \infty$

The virtual image is formed at an infinite distance to the left of the second lens. This acts as an object for the third lens.

$$\frac{1}{v_3} - \frac{1}{u_3} = \frac{1}{f_3}$$

or

$$\frac{1}{v_3} - \frac{1}{\infty} = \frac{1}{30} \text{ cm}$$

or $v_3 = 30\text{cm}$.

The final image is formed 30 cm to the right the third lens.

Ray diagrams and image formation for this combination of lenses are given in figure for example 11.4. Figure shows the ray diagram and image formation, for the first lens (Fig. (b)); for the first and second lens (Fig. (c)) and finally for the combination of all the three lenses (Fig. (d)).

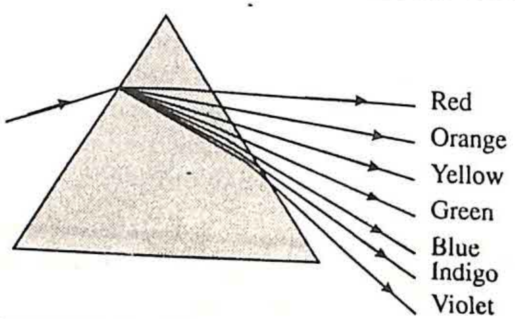
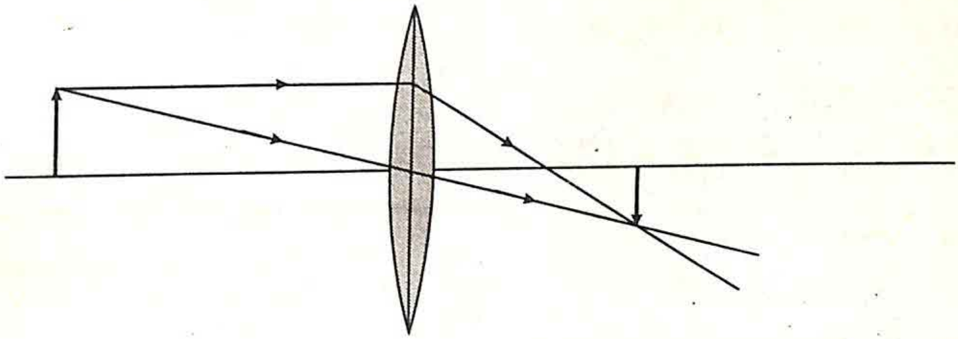


Figure 11.21: Dispersion of sunlight or white light passing through a glass prism. The relative deviation of different colours is shown highly exaggerated.

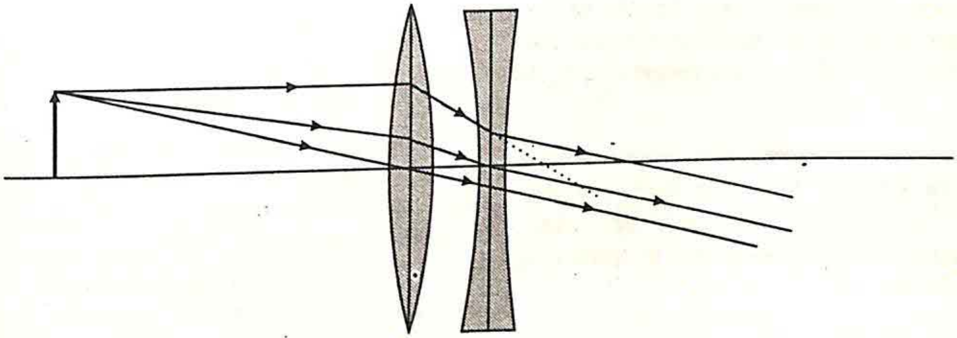
11.7 Dispersion, spectroscopes and spectra

11.7.1 Dispersion

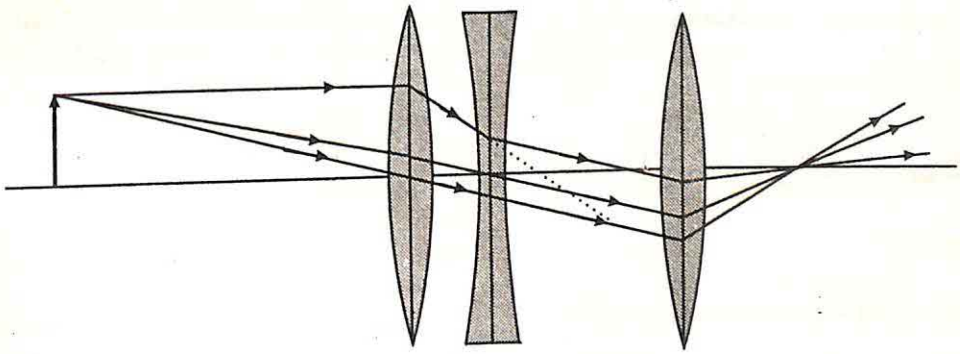
When a narrow beam of sunlight falls on a glass prism, the outcoming light is seen to consist of several colours. We say that sunlight has been *dispersed* by the prism to form a *solar spectrum*. There is a continuous variation of wavelength in this particular spectrum. It is conventional to say however that we see the seven colours violet, indigo, blue, green, yellow, orange and red in the



(b)



(c)



(d)

Figure for Example 11.4

Table 11.3: Refractive indices for different wavelengths

Colour	Wavelength (nm)	Crown glass	Flint glass
Violet	396.9	1.533	1.663
Blue	486.1	1.523	1.639
Yellow	589.3	1.571	1.627
Red	656.3	1.515	1.622

spectrum, the violet being the most deviated, the indigo next and so on and finally red, the least deviated colour (Fig. 11.21). The dispersion is due to the facts that sunlight is made up of light rays with a continuous range of wavelengths and that the refractive index of many materials including glass is different for different wavelengths (Table 11.3).

Example 11.5: Find the angle of dispersion between the red and violet colour produced by a flint glass prism of refracting angle of 60°.

Answer: We have the prism Eq. (11.20):

$$n = \frac{\sin[(A+D)/2]}{\sin(A/2)}$$

For the minimum deviation position

$$\sin\frac{A+D_{\text{red}}}{2} = n_{\text{red}} \sin\frac{A}{2}$$

$$\text{or} \quad D_{\text{red}} = 2\sin^{-1}\left(n_{\text{red}}\sin\frac{A}{2}\right) - A$$

Similarly

$$D_{\text{violet}} = 2\sin^{-1}\left(n_{\text{violet}}\sin\frac{A}{2}\right) - A$$

The dispersion between the red and violet colours is

$$D_{\text{violet}} - D_{\text{red}} = 2\sin^{-1}\left(n_{\text{violet}}\sin\frac{A}{2}\right)$$

$$-2\sin^{-1}(n_{\text{red}}\sin\frac{A}{2})$$

$$= 2\sin^{-1}(1.663 \times 1/2)$$

$$- 2\sin^{-1}(1.622 \times 1/2)$$

$$= 2 \times 56.25° - 2 \times 54.19° = 4.12°.$$

Light which has a single wavelength is called monochromatic light and many optical experiments need such a light source. For this either such special sources have to be used or a filter allowing only a particular wavelength has to be used if the source is polychromatic. In general many sources of light are polychromatic, giving light of several wavelengths.

11.7.2 Spectroscopes

These are instruments used to disperse polychromatic light namely, light containing different wavelengths and to study them. The spectroscope makes use of a component which produces differing deviation for different light wavelengths. It could be a plane diffraction grating (see Section 10.6.2) for which diffraction maxima occur at different angles for different wavelengths, or it could be a prism whose material has a refractive index that depends on wavelength so that light of different wavelengths is refracted differently. A common type of spectroscope is the prism spectroscope. It contains two assemblies and a prism in between them. The tube assembly A called a *collimator* has a narrow slit whose width can be adjusted by a screw. A light source whose spectrum is to be investigated is kept in front of the slit (S). The slit is arranged to be in the focal plane of the collimator lens (CL). Hence parallel light rays emerge from the collimator (Fig. 11.22).

These rays pass through the prism, get dispersed and enter the tube B called the

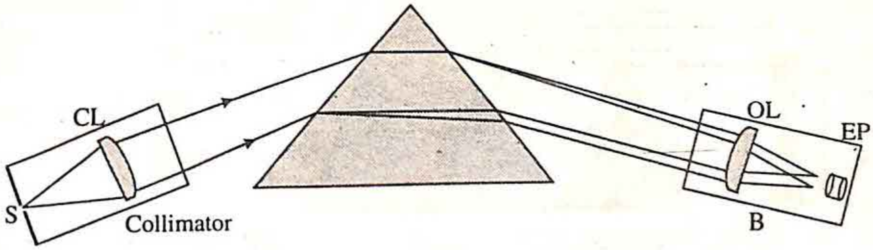


Figure 11.22: A schematic diagram of a prism spectroscope. The collimator, the prism and the telescope are shown.

telescope. The objective lens of the telescope forms a real image of the slit in a different direction for each wavelength of the light present. These images are either continuous or consist of separate lines depending on the nature of the source. The spectrum is magnified by the eye piece (EP). The prism is mounted on a rotating platform and the collimator and telescope also can be rotated about the common axis perpendicular to the prism table. A circular scale enables the positions of the collimator and telescope to be read and the angle of deviation to be measured. If the spectrometer is calibrated (for example by using standard wavelengths) the wavelength corresponding to any measured deviation can be obtained. In a spectrograph, a camera is substituted for the eye piece of the telescope. If a diffraction grating is substituted for the prism, one has a grating spectroscope, which gives wavelengths in terms of angle of diffraction and the spacing between grating lines.

11.7.3 Direct vision spectroscope

In a direct vision spectroscope, a number of thin prisms of crown and flint glasses are alternately placed in a small tube with refracting edges turned opposite to each other (Fig. 11.23). The angles are such that they produce dispersion without deviation of the mean ray.

There is a narrow adjustable slit S at one end of the tube, located at the focal plane of the achromatic converging lends L_1 (The word achromatic is explained in Section 11.9). Light enters through the slit and after passing through the lens and the prisms is collected by another converging lens L_2. This lens gives a spectrum at its focal plane and is observed through the eyepiece E.

The direct vision spectroscope is a pocket sized instrument meant for quick observation of many spectra.

11.7.4 Types of spectra

(i) Emission spectra

Spectra obtained from self-luminous bodies i.e. bodies emitting light by themselves are called *emission spectra*. There are three broad categories of emission spectra.

(a) *Continuous spectra*:- These are obtained for example from heated bodies and have a continuous range of wavelengths or colours.

(b) *Line spectra*:- These consist of narrow lines of distinct colours and dark intervals in between. (The 'lines' are

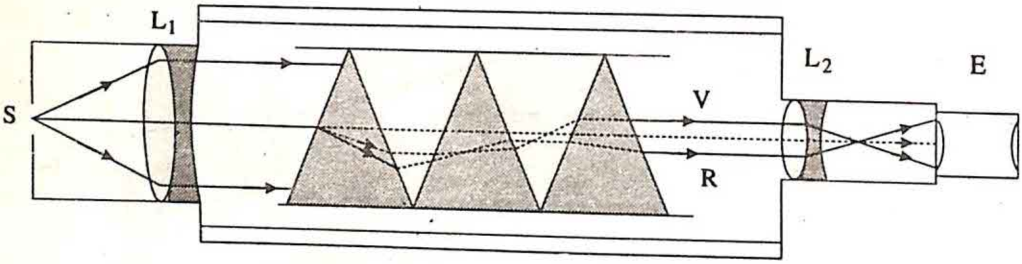


Figure 11.23: A direct vision spectroscope

just slit images observed in the spectroscope). Such line spectra are obtained from gases, for example. The sources can be made luminous by heating them or by striking an electric arc in the gas as in a discharge tube. Even solids can be made to give line spectra by burning them in a Bunsen burner and thereby converting them to the gaseous state. The line spectra help in identifying atoms in the constituent elements of the luminous object, since their wavelengths are characteristic of the atoms.

(c) *Band spectra:-* These consist of several bright bands interspersed with dark intervals and arise due to molecules. While free atoms (such as gases) give line spectra, atoms in combination (such as in molecules) give rise to band spectra.

(ii) Absorption spectra

When polychromatic light passes through a semitransparent substance (solid, liquid or gas), some of the particular colours of the incident radiation are absorbed. Then the transmitted radiation lacks several particular colours, with dark lines appearing in their place. Such a spectrum is called *absorption spectrum.* For example, the intense white light of a carbon arc lamp can be made to pass through sodium vapour. Then the continuous spectrum will have two dark lines corresponding to the two yellow lines of sodium. This is the absorption spectrum of sodium. Similarly the solar spectrum shows several dark lines in the continuous colour background. These are called the *Fraunhofer lines* and arise from the absorption of these particular colours of sunlight (i.e. light of the corresponding wavelength) by the atoms in solar atmosphere. The study of these Fraunhofer lines has enabled us to identify the abundance of hydrogen and helium in the solar atmosphere. As a matter of fact, this is how helium was discovered! A set of absorption lines was found to be prominent in the solar spectra,. These did not correspond to the spectral lines of any known element. This unknown element was called helium (from the Greek word *helios* for sun).

11.8 Light in nature

The play of light with things around us gives rise to several beautiful phenomena. The blue of the sky, white clouds, the red of sunrise and sunset, the rainbow, dappled sunlight as it filters through leaves of trees, and

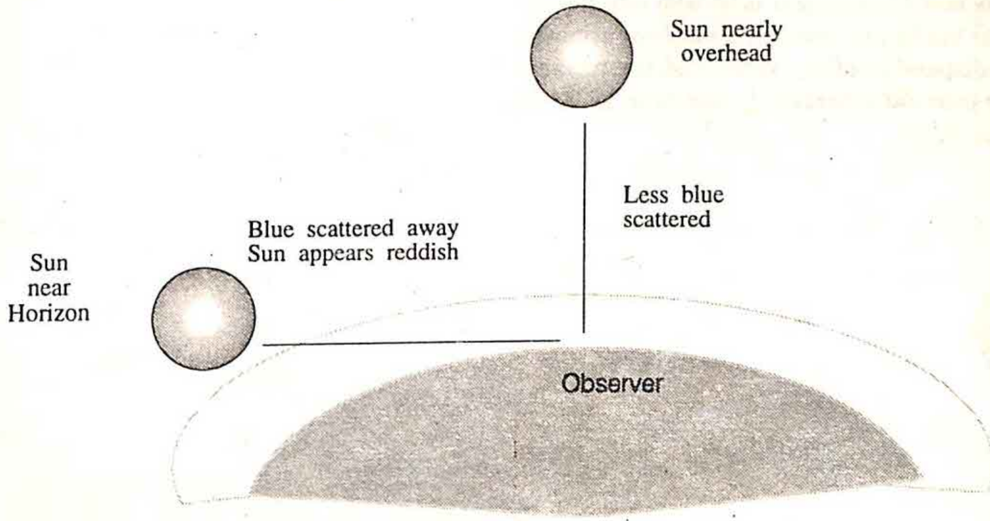


Figure 11.24: Absorption of sunlight at sunset and sunrise.

the iridiscent colours of some pearls, shells and wings of birds, are just a few of the natural wonders we are used to. We describe some of them here from the point of view of physics.

11.8.1 Scattering of Light

As sunlight travels through the earth's atmosphere, it gets *scattered* by the large number of molecules present. This scattering of sunlight is responsible for the colour of the sky, colour during sunrise and sunset, etc.. Light of shorter wavelength is scattered much more than light of larger wavelength. (The amount of scattering is inversely proportional to the fourth power of the wavelength. This scattering is known as Rayleigh scattering. We shall not try to explain it here). Hence the bluish colour of the sky predominates, since blue has a shorter wavelength than red and is scattered much more strongly.

Large particles like dust and water droplets present in the atmosphere do not have this selective scattering power. The quantity of relevance here is the relative size of the wavelength of light λ, and the scatterer (of typical size a, say). For $a << \lambda$, one has Rayleigh scattering which goes as $(1/\lambda)^4$. For $a >> \lambda$, i.e. large scattering objects (e.g. raindrops, large dust or ice particles) this is not true; all wavelengths are scattered nearly equally. Thus clouds are generally white.

At sunset or sunrise, sun's rays must pass through a larger atmospheric distance (Fig. 11.24). Most of the blue is scattered away. The sun may look almost reddish.

11.8.2 The Rainbow

The rainbow is another classic example of the dispersion of sunlight by the water drops in the atmosphere (Fig. 11.25). This is a phenomenon due to a combination of the refraction of sunlight by spherical water droplets behaving in a prism-like manner

and of total internal reflection. Fig. 11.25b shows how the sunlight is broken into its segments in the process and a rainbow appears. The dispersion of the violet and the red rays after internal reflection in the drop is shown in the figure.

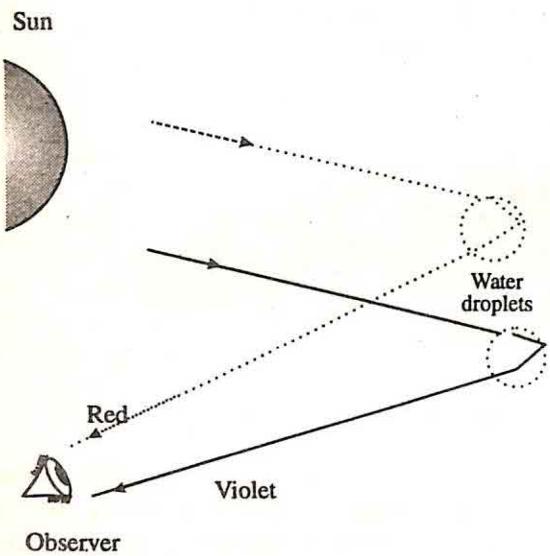


Figure 11.25(a)

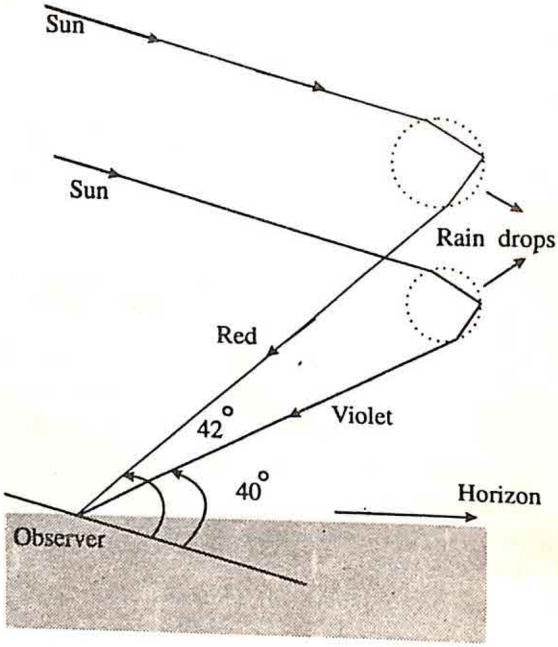


Figure 11.25(b)

The red rays emerge from the drops of water at one angle (42°) and the violet rays emerge at another angle (40°). The large number of water drops in the sky especially just after it has rained makes for a prominent rainbow. The parallel beam of sunlight getting dispersed at these angles produces a cone of rays at the observer as seen in the figure. The rainbow therefore appears circular as a arc for an observer on earth.

A secondary rainbow will sometimes be formed, with inverted colours. Here the light is reflected twice within the drop as shown in Fig. 11.25c. The secondary rainbow is fainter than the primary rainbow.

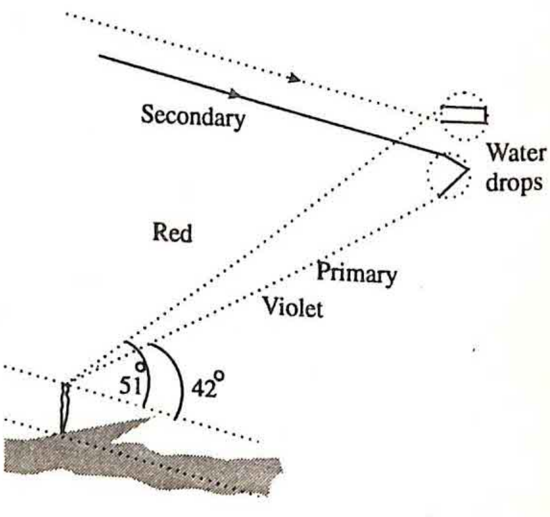


Figure 11.25(c)

Figure 11.25: Rainbow: (a) The sun, water drops, and the eye; (b) Total internal reflection and refraction of red and violet rays from raindrops and (c) Primary and secondary rainbows.

11.9 Optical defects in mirrors and lenses

Several kinds of optical defects can occur in mirrors and lenses. By this we mean that the image is not just a magnified or reduced version of the object. We shall name and describe some of the more important and common defects or aberrations. They are spherical aberration, coma, astigmatism, distortion and chromatic aberration.

11.9.1 Spherical aberration

If the reflecting or refracting surfaces are spherical, rays nearer to the edge of the lens or mirror are brought into focus at a shorter distance than those nearer the axis (see Fig. 11.26). Thus the image of either a point object or of an object at infinity is not a point. The envelope of the image rays forms a curve, called a caustic. (Fig. 11.26a for example). The intensity is a maximum on this curve, whose tip is the ideal image point. Spherical aberration is removed or reduced by using nonspherical surfaces. For example, a parabolic mirror brings parallel rays (rays parallel to its axis) to an *exact* focus at a point. (see Fig. 11.27). We cannot prove this geometrical property of a parabola here. One result is that all large telescopes use parabolic mirrors rather than spherical concave mirrors. (See Section 11.10 on optical instruments).

11.9.2 Coma

This is an optical defect in which a *point* object *off* the optic axis focusses not into a point image, but to a comet like surface (Fig. 11.26b).

11.9.3 Astigmatism

If we have an extended object, its image has different shapes at different distances! For

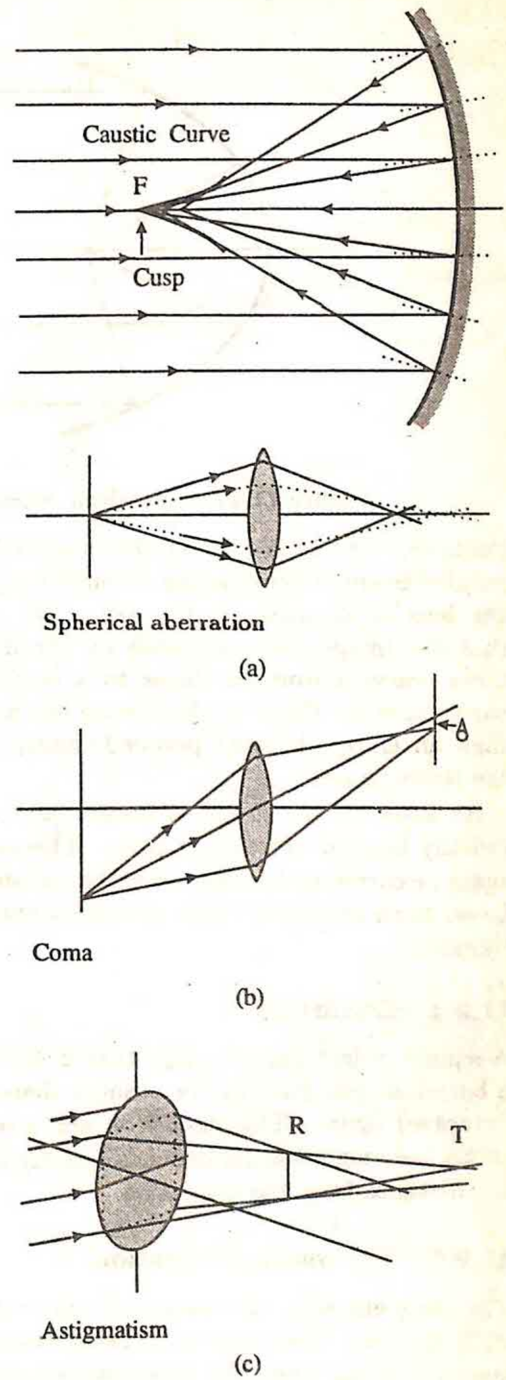


Figure 11.26: (a) Spherical aberration in a mirror and in a lens; (b) Coma and (c) Astigmatism.

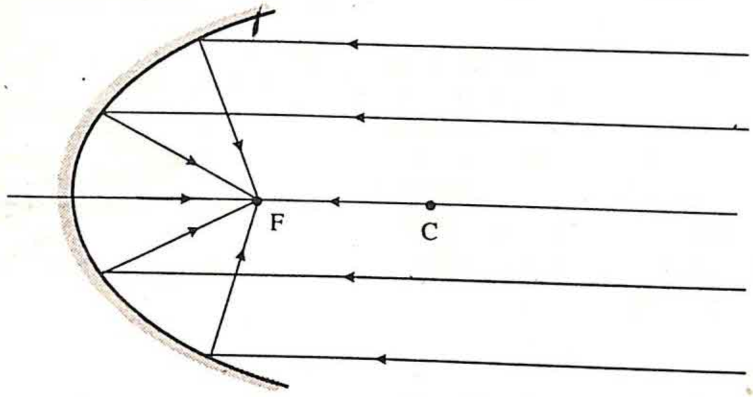


Figure 11.27: A parabolic concave mirror focusses parallel ray to a point.

example, Fig. 11.26c shows the image of a parallel beam of rays passing through a convex lens at an *angle* to the axis. We see that the image shape depends on the distance, varying from an ellipse to a vertical bar to a circle, then to a horizontal bar and then an ellipse directed perpendicularly to the previous one.

As a result, the image is tangentially or radially blurred in its outer parts. This can again be corrected by using nonspherical surfaces. Such corrected lenses are called anastigmatic.

11.9.4 Distortion

A square object has an image that is either a barrel shaped (convex) or cushion shaped (concave) figure. This distortion arises because the magnification depends slightly on the distance from the optic axis.

11.9.5 Chromatic aberration

This happens only in lenses, and has to do with the fact that rays of different wavelength coming from the *same* source, are bent differently. The focal length of a lens depends on the refractive index (Eq. (11.31)) which depends on the wavelength of light. This leads to coloured and fuzzy images of say white objects. This aberration is overcome by combining two or more lenses, made of materials with different dispersion, such that the extra deviation by one lens for one colour is compensated by the smaller deviation for the same colour due to the other lens. The lens combination has the *same* desired focal length for two or more wavelengths. Such a combination is called *achromatic*. A simple example of an *achromatic* doublet is shown in Fig. 11.28. The converging lens has chromatic aberration. The diverging lens is made of a different glass which has a much larger dispersion. This diverging lens has a larger (negative) focal length than the converging one, so that the two together act as a converging lens, bringing light of different wavelengths to focus at the same point. Modern lens design is very advanced. With lenses having several components (sometimes five or so) it is possible, in a modern camera, to form sharp images of objects of all sizes and colours (we discuss the camera in Section 11.10).

Finally, we note that there can be no chro-

matic aberration in mirrors. (The angle of reflection for the same angle of incidence does not depend on light wavelength).

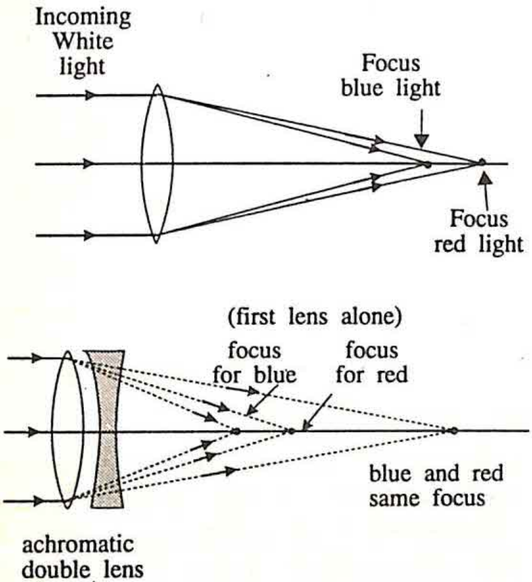


Figure 11.28: An achromatic doublet, consisting of a convex lens and a concave lens, the two have different dispersions and focal lengths differing in magnitude.

11.10 Optical instruments

A number of optical instruments are of daily use. The eye is the most common and essential, as well as the most unusual optical instrument. Starting with this, we briefly describe the principles of working of the camera, the microscope and the telescope. Finally, we go briefly into the question of the resolving power of optical instruments.

11.10.1 The eye

Fig. 11.29a shows the eye. Light enters the eye through a variable aperture, the pupil,

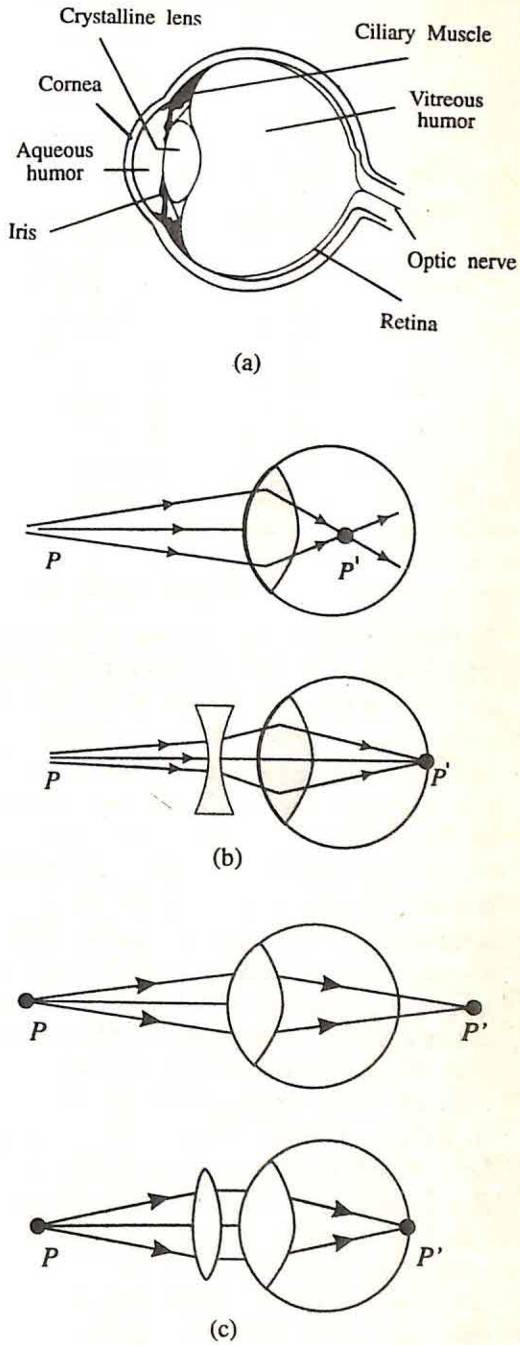


Figure 11.29: (a) The eye, showing some important parts; (b) The shortsighted, or myopic eye and (c) The farsighted, or hypermetropic eye.

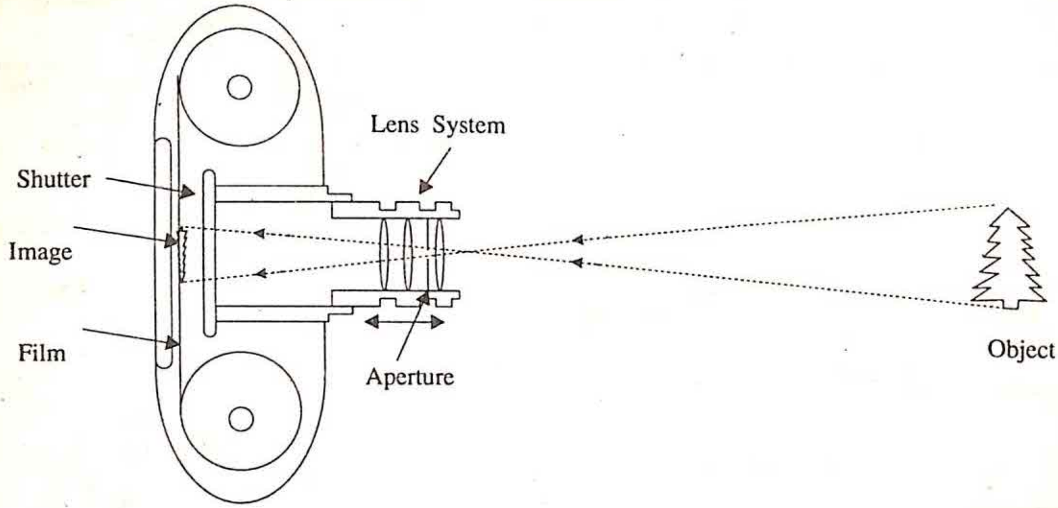


Figure 11.30: (a) The camera, with parts as shown.

and is focussed by the eye lens on the retina. The retina is a film of nerve fibres covering the curved back surface of the eye. The retina contains *rods* and *cones* which sense light (intensity as well as colour) and transmit electrical signals via the optic nerve to the brain which processes this information finally. The shape (curvature) and therefore the focal length of the lens can be modified somewhat by the *ciliary muscle*. For example, when the muscle is relaxed, the focal length is about 2.5 cm. When the object is brought closer to the eye, in order to maintain the same image lens distance ($\simeq$ 2.5 cm), the focal length of the eye lens needs to go up. This the ciliary muscles achieve by contracting, thus curving the lens further and increasing its focal length by about 0.1 cm. This property of the eye is called accommodation. If the object is too close to the eye, the lens cannot curve enough to focus the image on to the retina, and the image is blurred. The closest distance for which the lens can focus light on the retina is called the *least distance of distinct vision*, or *the near point*. This distance varies with age, decreasing with it because of the decreasing effectiveness of the ciliary muscle and the loss of flexibility of the lens. The near point may be as close as about 7 to 8 cm in a boy or girl ten years of age, and may increase to as much as 200 cm at 60 years of age. The standard value taken here is 25 cm. (Often the near point is given the symbol D).

A number of defects of vision are fairly common. For example, the lens may converge incident light to well before the retina. This is called *nearsightedness or myopia*. If we interpose a concave lens between the eye and the object, and the lens has the right diverging effect or negative focal length, the image will be focused on the retina. (Fig. 11.29b). Similarly, if the lens focuses at a distance farther than the retina, a convergent lens is needed. This defect is called *farsightedness* or *hypermetropia* (Fig. 11.29c).

Suppose the focal length needed is f cm. The power of the lens is $(100/f)$. The unit of lens power is dioptres. The sign of the

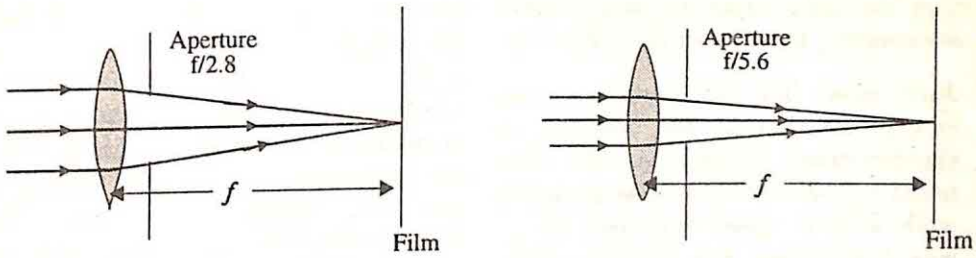


Figure 11.30: (b) Adjusting the aperture of the camera.

power indicates whether the lens is converging (+ power) or diverging (− power). For example, a person with a corrective lens of power −2.5 dioptres uses a concave lens of focal length (100/2.5) = 40 cm.

Another common defect vision is that of the cornea not being spherical, but having a larger curvature in one plane than in the other. This results in objects in one direction being well focused while those in a perpendicular direction are not. This is corrected by glasses with lenses of cylindrical rather than spherical shape. The defect is called astimagtism.

11.10.2 The camera

A photographic camera consists of a converging lens system at one end of a light proof box and light sensitive film at the other end (Fig. 11.30). A real inverted image of the object is formed on the film. The amount of light entering the film region can be adjusted by changing the aperture or opening of the lens.

A shutter is placed between the lens and the film. When a photograph is taken, the shutter opens and closes quickly thus exposing the film for a short time to light entering the camera through the lens.

For a good picture to be formed, the film should be exposed sufficiently in the short time when the shutter is open. So the following are the factors which the photographer tries to control.

(a) *Exposure time*: The film has to be exposed for a certain time for a clear image to be formed on the film. If the photograph is taken in bright sunlight, the exposure time can be small. In shade conditions or for indoor photography, the exposure time has to be larger. The normal exposure times used in cameras are

$$\frac{1}{500}\text{s.}, \frac{1}{250}\text{s.}, \frac{1}{125}\text{s.}, \frac{1}{60}\text{s.}, \frac{1}{30}\text{s, etc.}$$

(b) *Aperture of the camera*: This refers to the diameter of the circular opening through which the light passes into the camera. This is normally expressed as a fraction of f in f-numbers. Some apertures used in a camera are $(f/2)$, $(f/2.8)$, $(f/4)$, $(f/5.6)$, $(f/8)$, $(f/11)$, $(f/16)$, etc. An aperture of $(f/2.8)$ means the diameter d of the aperture is $f/2.8$. The area corresponding to aperture 5.6 is one fourth of the area corresponding to aperture 2.8. In order to have the same exposure of the film, the shutter will have to be kept open for four times the previous duration. You may also note that the numbers 2, 2.8, 4, 5.6, 8, 11, 16 are obtained by a successive multiplication by a factor of $\sqrt{2}$ i.e., these apertures correspond to shutter times (for get-

ting the same exposure) which differ successively by a factor of $(\sqrt{2})^2 = 2$.

(c) *Film speed*: The film speed is a measure of how quickly the film will be exposed when in use. A 'fast' film needs a relatively short time exposure while a 'slow' speed film needs somewhat longer time. Fast films are therefore used in poor lighting conditions and very slow films for still object photography.

(d) *Exposure meter*: Many cameras have built-in exposure meters. This has a light sensitive surface. Depending on the amount of light falling on it, a proportional amount of current flows in the meter. Using this, the photographer can adjust a suitable set of aperture and exposure times for correct exposure.

(e) *Depth of focus*: When the aperture is large, the image is not focused in a single plane, but gets spread out (see Section 11.9). The depth of focus indicates the range of the object distances over which the focusing is reasonably good.

It will be nice if you can practice and understand these details of photography. In addition, there are many further aids like wide angle and telephoto lenses, close-up adapters which make photography an exciting hobby.

11.10.3 The microscope

A simple magnifier or microscope is a converging lens of short focal length (Fig. 11.31). The lens is held near the object, one focal length away or less, and the eye is positioned close to the lens on the other side. The idea is to form an erect, magnified and virtual image of the object. This happens if the object is at a distance f (focal length of the lens) or less from the lens. If the object is at a distance f the image is at infinity. If the object is at a smaller distance, the image is virtual and closer than infinity. The closest comfortable distance for viewing is the near point (distance $D \simeq 25$ cm). Viewing at the near point causes some strain on the eye, so that often the image is at infinity, suitable for viewing by the relaxed eye. We show both cases, the first in Fig. 11.31(a), and the second in Figs 11.31(b) and (c).

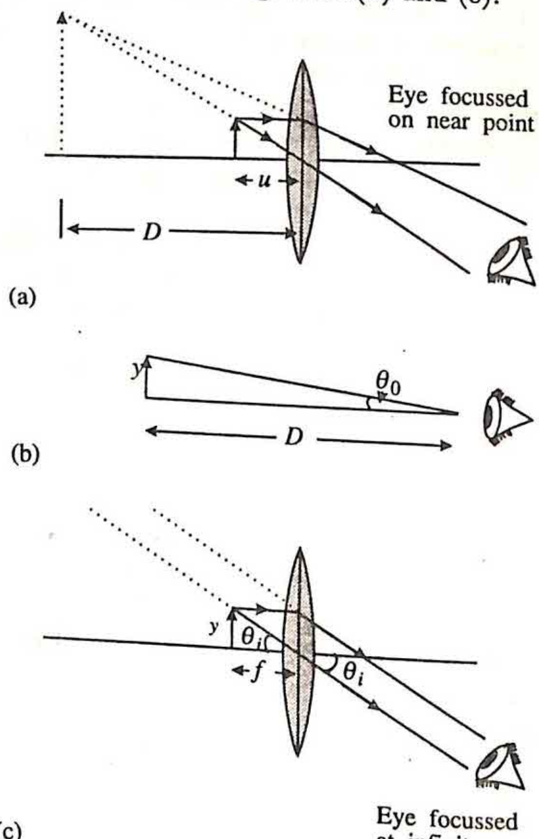


Figure 11.31: A simple microscope. (a) The magnifying lens is located such that the image is at the near point. (b) The object by itself, at the near point. (c) The object at the focal point of the lens; the image is at infinity.

The magnification when the image is

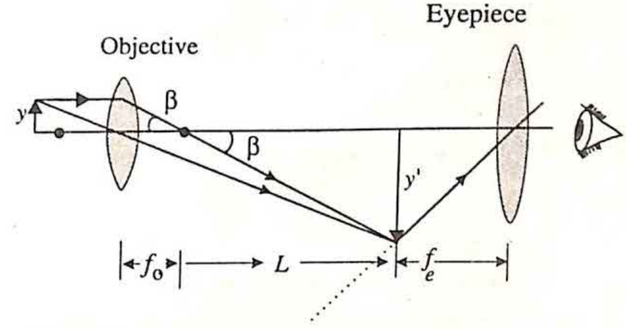


Figure 11.32: A compound microscope. The image is at infinity.

formed at the near point D is easily calculated. We have, the magnification M to be

$$M = \frac{v}{u} = v\left(\frac{1}{v} - \frac{1}{f}\right) = \left(1 - \frac{v}{f}\right)$$

Now according to our sign convention, v is negative, and is equal in magnitude to D. Thus the magnification is

$$M = \left(1 + \frac{D}{f}\right) \qquad (11.36)$$

For example, since D is about 25 cm, to have a magnification of six, one needs a convex lens of focal length $f = 5$ cm.

In the second case (the image at infinity), the image subtends a larger angle, forms a larger image on the retina, and so appears bigger. The magnification is equal to the *angular* magnification, or the ratio of the angles subtended by the object, and by the image, at the eye. Suppose the object has a length y. The maximum angle it can subtend, and be clearly visible (without a lens) is when it is at the near point, i.e. a distance D. The angle subtended then is

$$\theta_0 = (y/D). \qquad (11.37)$$

The angle subtended when the object is at the focal point of the magnifying lens so that it is imaged at infinity, is given by

$$\theta_i = (y/f) \qquad (11.38)$$

as is clear from Fig.(11.31c). The magnification is therefore

$$M = (\theta_i/\theta_0) = (D/f). \qquad (11.39)$$

This is one less than the magnification when the image is at the near point, Eq. (11.36), but the viewing is more comfortable, and the difference in magnification usually small. In subsequent discussions of optical instruments (microscope and telescope) we shall assume the image to be at infinity.

A simple microscope has a limited maximum magnification (≤ 10) for realistic focal lengths. For much larger magnifications, one uses two lenses, one compounding the effect of the other. This is known as the compound microscope. A schematic diagram is shown in Fig. 11.32. The lens nearest the object, called the objective, forms a real, inverted, magnified image of the object. This serves as the object for the second lens, the eyepiece or ocular, which functions essentially like a simple microscope or magnifier, producing an enlarged virtual final image. The first inverted image is thus near the focal point of the eye-piece, at a distance appropriate for final image formation at infinity, and a little closer for image formation at the near point. Clearly, the final image is inverted with respect to the original object.

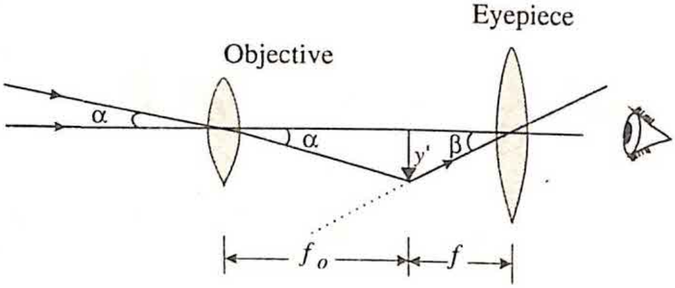


Figure 11.33: A refracting telescope.

We now obtain the magnification due to the compound microscope. The ray diagram of Fig. 11.32 shows that the (lateral) magnification due to the objective, namely (y'/y), equals

$$M_0 = \frac{y'}{y} = \frac{L}{f_0} \tag{11.40}$$

where we have used the result

$$\tan\beta = (y/f_0) = (y'/L).$$

Here y' is the size of the first image, the object size being y and the objective focal length being f_0. The first image is formed at the focal point of the eyepiece. The distance L, i.e. the distance between the second focal point of the objective and the first focal point of the eyepiece (focal length f_e) is called the tube length. This is because f_0 and f_e are rather small, and L is indeed nearly the length of the compound microscope tube.

Since the inverted first image is at the focal point of the eyepiece, we use the result from the discussion above for the simple microscope that the magnification M_e (angular) due to it is (see Eq. 11.39)

$$M_e = (D/f_e) \tag{11.41}$$

Thus the total magnification is

$$M = M_0 M_e = \left(\frac{L}{f_0}\right)\left(\frac{D}{f_e}\right) \tag{11.42}$$

Clearly, to achieve a large magnification of a *small* object (hence the name microscope), the objective and eyepiece should have a small focal lengths. It is impossible to make the focal length much smaller than 1 cm and avoid distortion.

For example, with an objective with $f_0 =$ 1.0 cm, and an eyepiece with focal length $f_e = 2.0$ cm, and a tube length of 20 cm, the magnification is

$$\begin{aligned} M = M_0 M_e &= \left(\frac{L}{f_0}\right)\left(\frac{D}{f_e}\right) \\ &= \frac{20}{1} \times \frac{25}{2} = 250. \end{aligned}$$

Various other factors such as illumination of the object contribute to the quality and visibility of the image. In modern microscopes, both the objective and the eyepiece consist of multicomponent lenses, to minimize various aberrations. This is specially necessary since focal lengths are small, i.e. lens surfaces are strongly curved. Magnifications of order 1000 are possible with very good design.

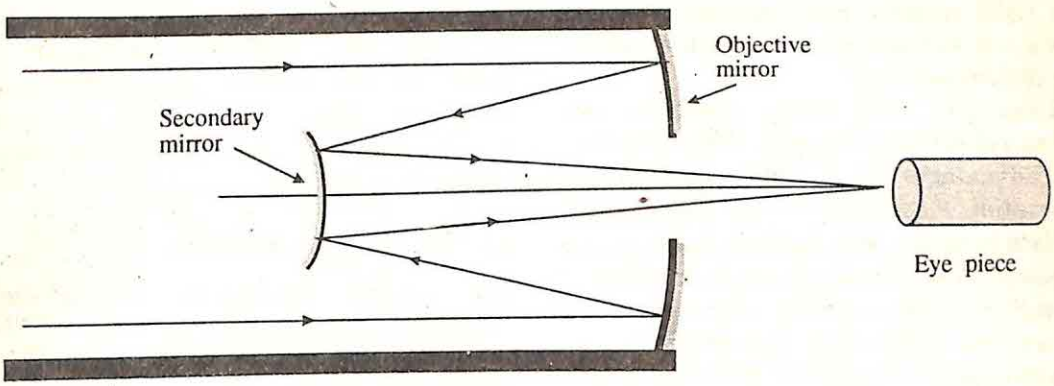


Figure 11.34: A reflecting telescope (Cassegrain).

11.10.4 Telescope

The telescope (Fig. 11.33) is used to provide angular magnification of distant objects. It also consists of an objective and an eyepiece. But here, the objective has a large focal length and a much larger aperture. Light from a distant object enters the objective and a real image is formed in the tube at the second focal point of the convex objective lens. The eyepiece magnifies this image producing a final inverted image. The magnifying power M corresponds to angular magnification. Angular magnification is defined as the ratio of the angle β subtended at the eye by the image to the angle α which the object subtends at the lens or the eye. Hence

$$M \approx \frac{\beta}{\alpha} \approx \frac{A'B'}{f_e}.\frac{f_0}{AB} = \frac{f_0}{f_e} \qquad (11.43)$$

Terrestrial telescopes, have in addition, a pair of inverting lenses to make the final image erect.

Refracting telescopes can be used for terrestrial and astronomical observations.

For example, consider a telescope whose objective has a focal length of 100cm and the eyepiece a focal length of 1 cm. The magnifying power of this telescope is

$$M = \frac{100}{1} = 100.$$

Let us consider a pair of stars of actual separation 12 min. The stars appear as though they are separated by an angle of 100 $\times$ 12 min = 20°.

The main considerations with an astronomical telescope are its light gathering power and its resolution or resolving power. The former clearly depends on the diameter of the objective (the amount of light admitted is proportional to the square of the diameter). With larger diameters, fainter (farther away?) objects can be observed. The resolving power, or the ability to observe as distinct two objects which are in very nearly the same direction, also depends on the diameter of the objective, as we shall see in the next subsection. So, the desirable aim in optical telescopes is to make them of large objective diameter. The largest lens objective in use (refracting telescope objective) has a diameter of 102 cm (It is at the Yerkes Observatory of the University of Chicago, at Williams Bay in Wisconsin, USA). Such big lenses (which have to be multicomponent to reduce various optical aberrations) tend to be very heavy and therefore difficult to support by their edges. It is also hard to make such lens combinations without aberrations.

For these reasons, most modern telescopes use a concave mirror rather than a lens for the objective.

Telescopes with mirror objectives are called *reflecting* telescopes. They have several advantages. First, there is no chromatic aberration in a mirror. Second, if a parabolic surface is chosen, spherical aberration is also removed. Mechanical support is much less of a problem since the mirror weighs much less than a lens of equivalent optical quality, and can be supported over its entire back surface, not just over its rim. One obvious problem with a reflecting telescope is that the objective mirror focuses light inside the telescope tube. One must have an eyepiece and the observer right there, obstructing some light (depending on the size of the observer cage). This is what is done in the very large 200" diameter Mt. Palomar telescope, California. The viewer sits near the focal point of the mirror, in a small cage. Another solution to the problem is to deflect the light being focused, by another mirror. One such arrangement, using a convex secondary mirror to focus the incident light, which now passes through a hole in the objective primary mirror, is shown in Fig.(11.34). This is known as a *Cassegrain* telescope, after its designer. It has a large focal length in a short telescope, and the convex secondary mirror slowly converges the reflected ray. There are also other types of reflecting arrangements such as the Newtonian reflector and the Schmidt reflector. The largest telescope in India is in Kavalur, Tamil Nadu. It is a 234 cm diameter reflecting telescope (Cassegrain). It was ground, polished, set up and is being used by the Indian Institute of Astrophysics, Bangalore.

The *prism binocular* is a double telescope using two sets of totally reflecting prisms. This makes the final image erect which is very desirable for observations on earth (terrestrial observation). It also effectively folds the optical path ABCDEF making for a shorter distance between the objective and the eyepiece (Fig. 11.35). Binoculars are much more compact and easier to use than a refracting telescope.

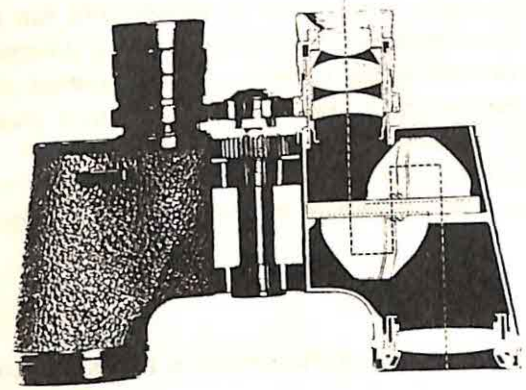
11.10.5 Resolving power

In discussing the optical instruments we have not considered the wave nature of light. This wave nature causes certain limitations. If a small hole is illuminated by a light source, then the diffraction pattern will be a set of circular fringes. These are alternately bright and dark with a central bright disk. If a lens is used to form an image, then the image of each point is similarly not a point but a diffraction disk. The size of the disk depends upon the aperture of the lens and the wavelength of the light used. (Think of the lens as a circular aperture). If we have two nearby points then their images may give rise to diffraction patterns overlapping on each other, making their identification as separate sources or points difficult. The resolving power of a lens is its ability to resolve two source points that are close to each other.

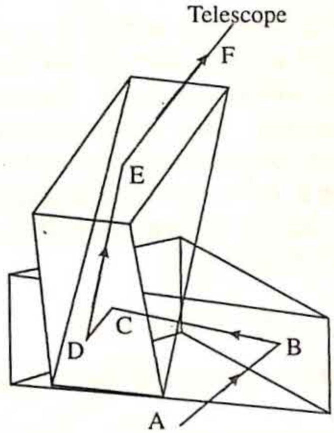
A rough idea of the limit of resolution can be obtained as follows. We have seen in Chapter 10 (Section 10.6.1) that a slit of width d produces, for parallel light incident on it, a diffraction pattern; the first maximum has an angular width of order (λ/d). The same is roughly true of a circular aperture of diameter d as well. If another source far away is in a direction less than (λ/d) radians away, then its image will overlap so strongly with that of the first source that they will be seen as one. So, the angular resolution due to the wave nature of light is

$$\Delta\theta \simeq (\lambda/d) \tag{11.44}$$

In case of the eye, two points can be seen distinctly if they susbtend at the eye an an-



(a)



(b)

Figure 11.35: (a) Binocular (b) Reflecting prisms in binocular.

gle equal to about one minute of arc. (This is on the assumption that the pupil of the eye is about 2 mm in diameter). The reciprocal of this angle is the resolving power.

The limit of the resolution of a microscope is determined by the least distance between two point objects which can be distinguished. This distance d is given by

$$d = \frac{\lambda}{2n \sin\theta} \qquad (11.45)$$

where λ is the wavelength of light used to illuminate the object, θ is the half angle of the cone of light from the point object and n the refractive index of the medium between the object and the objective. The expression $n \sin\theta$ is called numerical aperture, (for the eye, $n \sin\theta = 0.004$).

The resolving power of a telescope is defined as the reciprocal of the smallest angular separation between two distant objects whose images are separated in the telescope. It is given by

$$\Delta\theta = 1.22\,\frac{\lambda}{d} \qquad (11.46)$$

where λ is the wavelength of light, d is the diameter of the telescope objective and $\Delta\theta =$ angle subtended by the point object at the objective. To get a high resolving power, a telescope with large aperture objective has to be used. For example, the limit of the Kavalur telescope (diameter = 234cm) is approximately $(5500 \times 10^{-8})/234$ radians ~ 0.1 second of arc.

Summary

1. A source can radiate in a range of wavelengths, only a part of which (390 nm to 760 nm) is visible to the human eye. The sensitivity of the human eye is different for different colours. *Luminous flux* is defined as the luminous energy emitted per second. The unit of luminous flux is *lumen* (lm). Luminous intensity in a given direction is luminous flux in that direction per unit solid angle. Its unit is Candela. 1 cd is the luminous intensity in a given direction of a source that emits monochromtic radiation of frequency 540×10^{12} Hz and that has radiant intensity in that direction of (1/683) W/sr.

   $$1 \text{ lm} = 1 \text{ cd sr}$$

   Illuminance is luminous flux per unit area. The unit of illuminance is called the lux:

   $$1 \text{ lux} = 1 \text{ lm m}^{-2}.$$

2. The speed of light c was estimated roughly by Romer from observations at the times of eclipse of the moons of Jupiter. More accurate measurements of c were made by the terrestrial experiments due to Fizeau, and later by Michelson. The speed of light in free space is a fundamental constant whose value is now taken to be

   $$c = 299792458 \text{ ms}^{-1}$$

3. *New Cartesian sign convention*: Distance measured in the same direction as the incident light are positive; those in the opposite direction are negative. All distances are measured from the centre of the mirror/lens on the optic axis.

   Mirror equation:

   $$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$

   where u and v are object and image distances, respectively and f is the focal length of the mirror. 'f' is approximately half the radius of curvature R. $f < 0$ for concave mirror; $f > 0$ for convex mirror.

4. *The critical angle of incidence* from a denser to rarer medium is given by

   $$\sin i_c = \frac{n_{\text{rarer}}}{n_{\text{denser}}}$$

   For $i > i_c$, total internal reflection occurs. Examples: multiple internal reflections in diamond ($i_c = 24.41^o$), totally reflecting prisms, mirage.

   Optical fibres consist of glass fibres coated with a thin layer of material of *lower* refractive index. Light incident at one end comes out after multiple internal reflections, even if the fibre is bent.

5. For a prism of the angle A, refractive index n_2 placed in a medium of refractive index n_1,

$$\frac{n_2}{n_1} = \frac{\sin\left(\frac{A+D}{2}\right)}{\sin\frac{A}{2}}$$

where D is the angle of minimum deviation.

6. *Refraction through a spherical interface*

(from medium 1 to 2 of refractive index n_1 and n_2, respectively)

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R}$$

Thin lens formula:

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

where $\frac{n_1}{f} = (n_2 - n_1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right]$,

R_1 and R_2 being the radii of curvature of the lens surfaces.
$f > 0$ for convex lens; $f < 0$ for concave lens.

7. *Dispersion* is the splitting of light into its constitutent colours. A spectroscope is an instrument to disperse polychromatic light and measure different wavelengths. It consists of a collimator, a prism and a telescope. A direct vision spectroscope produces dispersion without deviation of the mean ray. The rainbow is an example of dispersion of sunlight by the water drops in the atmosphere.

8. *Emission spectra* are of three types: *Continuous Spectra* obtained from heated bodies; *Line spectra* (consisting of narrow distinct colours) obtained, for example, by discharge through gases; *band spectra*: bright bands interspersed with dark intervals. Free atoms give their characteristic line spectra; atoms in combination such as molecules give band spectra.

9. *Optical Defects of Lenses and Mirrors:* The image of a point object produced by lenses and curved mirrors is not a point, for many reasons. These optical defects called aberrations, are sperical aberration, coma, astigmatism, and chromatic aberration, among others. Chromatic aberration is due to light of different wavelengths having differing focal lengths in a lens. The optical defects can be minimized by using nonspherical surfaces (eg. sperical aberration) as well as combinations of lenses (chromatic aberration).

10. *The Eye:* The eye has a convex length of focal length about 2.5 cm. This focal length can be varied somewhat so that the image is always on the retina. This ability of the eye is called accommodation. If the image is focussed before the retina, (myopia) a concave corrective lens is needed; if the image is focussed beyond the retina, (hypermetropia) a convex corrective lens is needed.

11. In a *camera*, the aperture size and exposure time together determine the total amount of light received. An aperture $f/11$ means the diameter of aperture is $f/11$. The sequence $f/2$, $f/2.8$, $f/4$, $f/5.6$, $f/8$, $f/11$ etc. is such that the exposure times required to get the *same* amount of light are in the ratio 1:2:4:8:16:32.

12. *Magnifying power of a simple microscope* has a magnitude M given by

$$M = 1 + (D/f)$$

where $D = 25$ cm is the least distance of distinct vision and $f(> 0)$ is the focal length of convex lens. For a compound microscope, the magnifying power is given by

$$M = M_e \times M_0$$

where $M_e = 1 + (D/f_e)$ is the magnification due to the eyepiece and M_0 is the magnification produced by the objective. *Approximately,*

$$M = \frac{L}{f_0} \times \frac{D}{f_e}$$

f_0 and f_e are the focal lengths of the objective and eye-piece respectively, and L is the distance between them.

13. *Magnifying power M of a telescope* is the ratio of the angle β subtended at the eye by the image to the angle α subtended at the eye by the object.

$$M = \frac{\beta}{\alpha} = \frac{f_0}{f_e}$$

where f_0 and f_e are the focal lengths of the objective and eye-piece, respectively.

14. *The resolution of a microscope* i.e., the least distance d between two objects which can be distinguished is given by

$$d = \frac{\lambda}{2n \sin\theta}$$

where λ is the wavelength of light, θ is the half angle of the cone of light from the point object, and n is the refractive index of the medium between the object and the objective. The resolving power of a telescope is the reciprocal of the smallest angular separation $\Delta\theta$ between two distant objects with distinct images.

$$\Delta\theta = \frac{1.22\,\lambda}{a}$$

where a is the diameter of the telescope objective.

Exercises

11.1 A lamp placed 70.0 cm from a screen on one side produces the same illumination as a standard 60 cd lamp placed 105 cm away on the other side of the screen. What is the luminous intensity of the first lamp? [Principle of photometer].

11.2 To print a photograph from a negative, the time of exposure to light from a lamp placed 0.50 m away is 2.5 s. How much exposure time is required if the lamp is placed 1.0 m away?

11.3 An astronomer observes the eclipse of a moon of Jupiter at two different times of the year (about 200 days apart), first when the Earth and the Jupiter are on the same side of the sun (and in line with it) and second when they are on the opposite sides of the sun. The second eclipse takes place about 990 s later than expected on the basis of the geometry of the orbits. Estimate the speed of light given that the orbital radius of the Earth is about 1.5×10^{11} m (Romer's method).

11.4 In Fizeau's method for the determination of c, a wheel with 540 teeth is used and the source of light disappears from view when the period of revolution of the wheel is decreased to 0.072 s. Estimate the value of c, if the far mirror is at a distance of 10 km from the wheel.

11.5 In Michelson's rotating (octagonal) prism method for the determination of c, the distance between the rotating prism and the distant mirror is measured to be 45.0 km. A minimum speed of 416.7 rev/s for the rotation of the prism is needed to view the source in the same position as when the prism is at rest. Determine the value of c.

11.6 Derive the mirror formula connecting object distance, image distance and focal length. Specify your sign convention.

11.7 A small candle 2.5 cm in size is placed 27 cm in front of a concave mirror of radius of curvature 36 cm. At what distance from the mirror should a screen be placed in order to receive a sharp image? Describe the nature and size of the image. If the candle is moved closer to the mirror, how would the screen have to be moved?

11.8 A 4.5 cm needle is placed 12 cm away from a convex mirror of focal length 15 cm. Give the location of the image and the magnification. Describe what happens as the needle is moved farther from the mirror.

11.9 A square wire of side 3.0 cm is placed 25 cm away from a concave mirror of focal length 10 cm. What is the area enclosed by the image of the wire? (The centre of the wire is on the axis of the mirror, with its two sides normal to the axis.)

11.10 A tank is filled with water to a

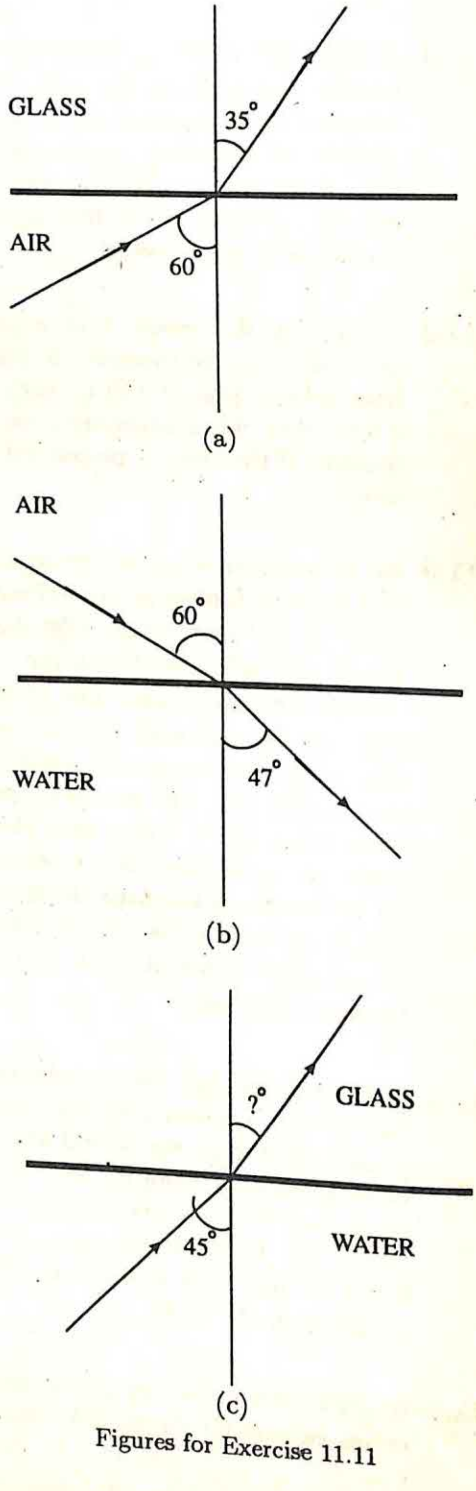
height of 12.5 cm. The apparent depth of a needle lying at the bottom of the tank is measured by a microscope to be 9.4 cm. What is the refractive index of water? If water is replaced by a liquid of refractive index 1.63 upto the same height, by what distance would the microscope have to be moved to focus on the needle again?

11.11 Figures (a) and (b) show refraction of an incident ray in air at 60° with the normal to a glass-air and water-air interface, respectively. Predict the angle of refraction of an incident ray in water at 45° with the normal to a water-glass interface (Fig. (c)).

11.12 A small bulb is placed at the bottom of tank containing water to a depth of 80 cm. What is the area of the surface of water through which light from the bulb can emerge out? Refractive index of water is 1.33. [Consider the bulb to be a point source].

11.13 A prism is made of glass of unknown refractive index. A parallel beam of light is incident on a face of the prism. By rotating the prism, the minimum angle of deviation is measured to be 40°. What is the refractive index of the prism? If the prism is placed in water (refractive index 1.33), predict the new minimum angle of deviation of a parallel beam of light. The refracting angle of the prism is 60°.

11.14 For refraction at a spherical surface, derive a formula connecting object



Figures for Exercise 11.11

distance, image distance and radius of curvature of the surface. Specify your sign convention.

11.15 Use the formula obtained in 11.14 above to derive the thin lens formula

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}.$$

where the symbols have the usual meanings. Specify the sign convention adopted in this formula.

11.16 A needle placed 45 cm from a lens forms an image on a screen placed 90 cm on the other side of the lens. Identify the type of the lens and determine its focal length. What is the size of the image if the size of the needle is 5.0 cm?

11.17 Double-convex lenses are to be manufactured from a glass of refractive index 1.55, with both faces of the same radius of curvature. What is the radius of curvature required if the focal length is to be 20 cm?

11.18 A beam of light converges to a point P. A lens is placed in the path of the convergent beam 12 cm from P. At what point does the beam converge if the lens is (a) a convex lens of focal length 20 cm, (b) a concave lens of focal length 16 cm?

11.19 An object of size 3.0 cm is placed 14 cm in front of a concave lens of focal length 21 cm. Describe the image produced by the lens. What happens if the object is moved farther from the lens?

11.20 (a) Derive the relation

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

where f_1 and f_2 are the focal lengths of two (thin) lenses and f is the focal length of the combination of the two lenses in contact.

(b) What is the focal length of a convex lens of focal length 30 cm in contact with a concave lens of focal length 20 cm? Is the system a converging or a diverging lens? Ignore thickness of the lenses.

11.21 (a) What is meant by 'magnifying power of a microscope' ?

(b) A thin convex lens of focal length 5 cm is used as a simple microscope by a person with normal near point (25 cm). What is the magnifying power of the microscope?

11.22 A compound microscope consists of an objective lens of focal length 2.0 cm and an eyepiece of focal length 6.25 cm separated by a distance of 15 cm. How far from the objective should an object be placed in order to obtain the final image at (a) the least distance of distinct vision (25 cm), (b) infinity? What is the magnifying power of the microscope in each case?

11.23 A person with a normal near point (25cm) using a compound microscope with objective of focal length 8.0 mm and an eyepiece of focal length 2.5 cm can bring an object placed 9.0 mm from the objective in

sharp focus. What is the separation between the two lenses? How much is the magnifying power of the microscope?

11.24 (a) What is meant by 'magnifying power of a telescope'?

(b) A small telescope has an objective lens of focal length 144 cm and an eyepiece of focal length 6.0 cm. What is the magnifying power of the telescope? What is the separation between the objective and the eyepiece?

11.25 (a) A giant refracting telescope at an observatory has an objective lens of focal length 15 m. If an eyepiece of focal length 1.0 cm is used, what is the angular magnification of the telescope?

(b) If this telescope is used to view the moon, what is the diameter of the image of the moon formed by the objective lens? The diameter of the moon is 3.48×10^6 m, and the radius of lunar orbit is 3.8×10^8 m.

11.26 A telescope has an objective of diameter 60 cm. The focal lengths of the objective and eye-piece are 2.0 m and 1.0 cm, respectively. The telescope is directed to view two distant almost point sources of light (e.g. two stars of a binary). The sources are at roughly the same distance ($= 10^4$ light years) along the line of sight but are separated transverse to the line of sight by a distance of 10^{10} m. Will the telescope resolve the two objects i.e. will it see two distinct stars?

11.27 (a) An object is placed between two plane mirrors inclined at 60° to each other. How many images do you expect to see?

(b) An object is placed between two plane parallel mirrors. Why do the distant images get fainter and fainter?

(c) Why are mirrors used in searchlights parabolic and not concave spherical?

(d) If you were driving a car, what type of mirror would you prefer to use for observing traffic at your back?

11.28 (a) A concave mirror and a convex lens are held in water. What change, if any, do you expect to find in the focal length of either?

(b) On a hot summer day in a desert, one sees the reflected image of distant parts of the sky. (This is sometimes mistaken by the observer to be the reflection of the sky in some distant lake of water. This illusion is called a mirage). Explain.

(c) What is the twinkling effect of starlight due to?

(d) Watching the sunset on a beach, one can see the sun for several minutes after it has 'actually set', Explain.

11.29 (a) People usually prefer light-coloured dresses during summer and dark dresses during winter. Why?

(b) How would a blue object appear under sodium lamp light?

(c) What does a welder protect against when he wears a mask?

(d) Explain why the sky is blue, and the sun appears red at sunset.

11.30 (a) What is the adjustment needed in a camera to take pictures of objects at different distances?

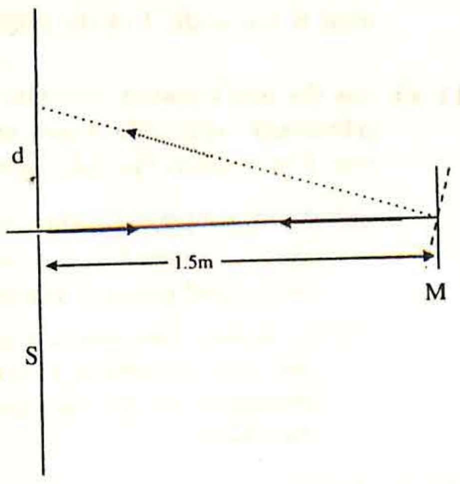
(b) What does the (adjustable) *f*-number of a camera signify?

What does one mean by saying that the aperature is f/11. Why are aperatures labelled as $f/2$, $f/2.8$, $f/4$, $f/5.6$, $f/8$, $f/11$ etc?

(c) What is the function of the camera shutter?

Additional Exercises

11.31 Light incident normally on a plane mirror attached to a galvanometer coil retraces backwards as shown. A current in the coil produces a deflection of 3.5° of the mirror. What is the displacement of the reflected spot of light on a screen placed 1.5 m away?



11.32 A boy 1.50 m tall with his eye level at 1.38 m stands before a mirror fixed on a wall. Indicate by means of a ray diagram how the mirror should be positioned so that he can view himself fully. What should be the minimum length of the mirror? Does the answer depend on the eye level?

11.33 Light incident on a rotating mirror M is reflected to a fixed mirror N placed 22.5 km away from M. The fixed mirror reflects it back to M (along the same path) which in turn reflects the light again along a direction that makes an angle of 27° with the incident direction. What is the speed of rotation of the mirror if the speed of light is 3.0×10^8 ms^{-1}? (Principle of Foucault's method for the determination of c).

11.34 Use the mirror equation to deduce that

(a) an object placed between f and $2f$ of a concave mirror produces a real image beyond $2f$

(b) a convex mirror always produces a virtual image independent of the location of the object

(c) the virtual image produced by a convex mirror is always diminished in size and is located between the focus and the pole.

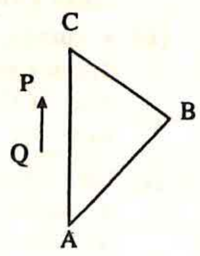
(d) an object placed between the pole and focus of a concave mirror produces a virtual and enlarged image.

[*Note:* This exercise helps you deduce algebraically properties of images that one obtains from explicit ray diagrams].

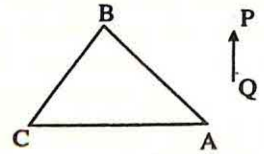
11.35 A small pin fixed on a table top is viewed from above from a distance of 50 cm. By what distance would the pin appear to be raised if it is viewed from the same point through a 15 cm thick glass slab held parallel to the table? Refractive index of glass = 1.5. Does the answer depend on the location of the slab?

11.36 The bottom of a container is a 4.0 cm thick glass (n = 1.5) slab. The container contains two immiscible liquids A and B of depths 6.0 cm and 8.0 cm, respectively. What is the apparent position of a scratch on the outer surface of the bottom of the glass slab when viewed through the container? Refractive indices of A and B are 1.4 and 1.3, respectively.

11.37 A right-angle prism is placed before an object in the two positions shown. The prism is made of crown glass with critical angle equal to 41°.



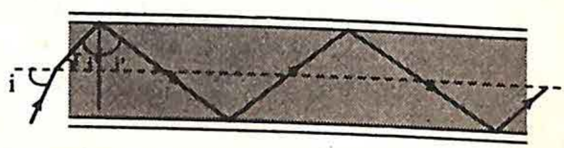
(a)



(b)

Trace the paths of two rays from P and Q normal to the hypotenuse in (a), and parallel to the hypotenuse in (b).

11.38 (a) The figure shows a cross-section of a 'light pipe' made of a glass fibre of refractive index 1.68. The outer covering of the pipe is made of a material of refractive index 1.44. What is the range of the angles of the incident rays with the axis of the pipe for which total reflections inside the pipe take place as shown.



(b) What is the answer if there is no outer covering of the pipe?

11.39 Parallel light from the collimator of a spectrometer is incident on the two faces of a prism which make the refracting angle A of the prism. The image of the collimator slit is observed in two different positions of the telescope of the spectrometer. If the angle of rotation of the telescope between the two positions is 144°, what is the angle A of the prism?

11.40 Use the lens equation to deduce algebraically what you know otherwise from explicit ray diagrams:

(a) An object placed within the focus of a convex lens produces a virtual and enlarged image.

(b) A concave lens produces a virtual and diminished image independent of the location of the object.

11.41 Answer the following questions:

(a) A man holding a lighted candle

in front of a thick glass mirror and viewing it obliquely sees a number of images of the candle. What is the origin of these multiple images?

(b) You read a newspaper because of the light that it reflects. Then why do you not see even a faint image of yourself in the newspaper?

(c) You have learnt that plane and convex mirrors produce virtual images of objects, can they produce real images under some circumstances? Explain.

(d) The wall of a room is covered with a perfect plane, mirror, and two movie films are made, one recording the movement of a man and the other of his mirror image. From viewing the films later can an outsider tell which is which?

11.42 Answer the following questions:

(a) A virtual image, we always say, cannot be caught on a screen. Yet when we 'see' a virtual image, we are obviously bringing it on to the 'screen' (i.e. the retina) of our eye. Is there a contradiction?

(b) To a fish under water viewing obliquely a fisherman standing on the bank of a lake, does the man look taller or shorter than what he actually is?

(c) Does the apparent depth of a tank of water change if viewed obliquely? If so, does the apparent depth increase or decrease?

(d) The refractive index of diamond is much greater than that of ordinary glass. Is this fact of some use to a diamond-cutter?

11.43 The image of a small electric bulb fixed on the wall of a room is to be obtained on the opposite wall 3 m away by means of a large convex lens. What is the maximum possible focal length of the lens required for the purpose?

11.44 (a) A screen is placed 90 cm from an object. The image of the object on the screen is formed by a convex lens at two different locations separated by 20 cm. Determine the focal length of the lens.

(b) Suppose the object in (a) above is an illuminated slit in a collimator tube so that it is hard to measure the slit size and its distance from the screen. Using a convex lens one obtains a sharp image of the slit on a screen. The image size is measured to be 4.6 cm. The lens is displaced away from the slit and at a certain location, another sharp image of size 1.7 cm is obtained. Determine the size of the slit.

11.45 (a) Determine the 'effective focal length' of the combination of the

two lenses in 11.20(b) if they are placed 8.0 cm apart with their principal axes coincident. Does the answer depend on which side a beam of parallel light is incident? Is the notion of effective focal length of this system useful at all?

(b) An object 1.5 cm in size is placed on the side of the convex lens in the above arrangements. The distance between the object and the convex lens is 40 cm. Determine the magnification produced by the two-lens system, and the size of the image.

11.46 At what angle should a ray of light be incident on the face of a prism of refracting angle 60° so that it just suffers total internal reflection at the other face? The refractive index of the prism is 1.524.

11.47 The figure shows an equiconvex lens (of refractive index 1.50) in contact with a liquid layer on top of a plane mirror. A small needle with its tip on the principal axis is moved along the axis until its inverted image is found at the position of the needle. The distance of the needle from the lens is measured to be 45.0 cm. The liquid is removed and the experiment is repeated. The new distance is measured to be 30.0 cm. What is the refractive index of the liquid?

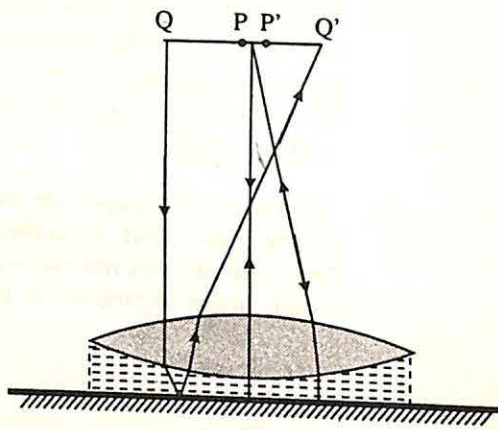


Figure for Exercise 11.47

11.48 Using a spectrometer, the following data are obtained for a crown glass prism and a flint glass prism.

Crown glass prism:

Angle of the Prism, A = 72.0°

Minimum deviation angle:

$$\delta_b = 54.6°,$$
$$\delta_r = 53.0°,$$
$$\delta_y = 54.0°$$

Flint glass prism:

A = 60.0°

$$\delta_b = 52.8°,$$
$$\delta_r = 50.6°,$$
$$\delta_y = 51.9°,$$

b, r and y refer to particular wavelengths in the blue, red and yellow bands. Compare the dispersive powers of the two varieties of glass prisms.

11.49 You are given prisms made of crown glass and flint glass with a wide variety of angles. Suggest a combination of prisms which will (a) deviate a pencil of white light without much dispersion, (b) disperse (and displace) a pencil of white light without much deviation. Use quali-

tatively the answer to 11.48.

11.50 A convex lens made of a variety of glass of high dispersive power has focal length of 15 cm. A parallel beam of white light is incident on one side of the lens and screen is placed on the other side. Describe the chromatic aberration of the lens, i.e., describe the colours on the spot focussed on the screen as the screen is moved away from the lens.

11.51 (a) A combination of two thin lenses in contact is to be made which has the same length for blue and red light. [Such a combination is known as an 'achromatic doublet']. Show that the ratio of their focal lengths (for yellow light) must be equal in magnitude and opposite in sign to the ratio of the dispersive powers of the materials of the two lenses.

(b) Use the results in (a) to suggest a way of removing chromatic aberration of the lens in 11.50 which is made of flint glass. You are given convex and concave lenses (made of crown glass) of various focal lengths. The ratio of the dispersive powers of flint glass to crown glass is about 1.5.

11.52 You are given a double-convex lens made of crown glass with each surface of radius of curvature 15 cm. A flint glass lens is grafted on to one of the surfaces of this lens. What is the radius of curvature of the second surface of the flint glass lens for the combination to be an 'achromatic doublet' for blue and red light? Data on refractive indices required may be obtained from answer to 11.48.

11.53 Answer the following questions:

(a) Do materials always have the same colour whether viewed by reflected light or through transmitted light?

(b) What colour do you observe when white light passes through a blue and yellow filter?

11.54 For a normal eye, the far point is at infinity and the near point of distinct vision is about 25 cm in front of the eye. The cornea of the eye provides a converging power of about 40 dioptres, and the least converging power of the eyelens behind the cornea is about 20 dioptres. From this rough data estimate the range of accommodation (i.e. the range of converging power of the eyelens) of a normal eye.

11.55 Does short-sightedness (myopia) or long-sightedness (hyperopia) imply *necessarily* that the eye has partially lost its ability of accommodation? If not, what might cause these defects of vision?

11.56 (a) The far point of a myopic person is 80 cm in front of the eye. What is the power of the lens required to enable him to see very distant objects clearly?

(b) In what way does the corrective lens help the person above? Does the lens magnify very distant objects? Explain carefully.

(c) The person above prefers to remove his spectacles while reading a book. Explain why?

11.57 (a) The near point of a hyperopic person is 75 cm from the eye. What is the power of the lens required to enable him to read clearly a book held at 25 cm from the eye?

(b) In what way does the corrective lens help the person above? Does the lens magnify objects held near the eye?

(c) The person above prefers to remove his spectacles while looking at the sky. Explain why?

11.58 (a) Suppose the person in 11.56 uses spectacles of power −0.80 dioptres, how far can he see clearly?

(b) If the person in 11.57 uses spectacles of power +1.0 dioptres, what is the nearest distance of distinct vision for him?

11.59 A myopic person has been using spectacles of power −1.0 dioptre for distant vision. During old age he also needs to use separate reading glass of power +2.0 dioptres. Explain what may have happened.

11.60 A person looking at a mesh of crossed wires is able to see the vertical wires more distinctly than the horizontal wires. What is this defect due to? How is such a defect of vision corrected?

11.61 A man with normal near point (25 cm) reads a book with small print using a magnifying glass: a thin convex lens of focal length 5 cm.

(a) What is the closest and the farthest distance at which he can read the book when viewing through the magnifying glass?

(b) What is the maximum and the minimum angular magnification (magnifying power) possible using the above simple microscope?

11.62 A figure divided into squares each of size 1 mm^2 is being viewed at a distance of 9 cm through a magnifying glass (a converging lens of focal length 10 cm) held close to the eye.

(a) What is the magnification (image size/object size) produced by the lens? How much is the area of each square in the virtual image?

(b) What is the angular magnification (magnifying power) of the lens?

(c) Is the magnification in (i) equal to the magnifying power in (ii). Explain.

11.63 (a) At what distance should the lens be held from the figure in 11.62 in order to view the squares distinctly with the maximum possible magnifying power?

(b) What is the magnification (image size/object size) in this case?

(c) Is the magnification equal to the magnifying power in this case? Explain.

11.64 What should be the distance between the object in 11.62 and the

magnifying glass if the virtual image of each square in the figure is to have an area of 6.25 mm^2. Would you be able to see the squares distinctly with your eyes very close to the magnifier?

Note: Exercises 11.62 to 11.64 will help you clearly understand the difference between magnification in absolute size and the angular magnification (or magnifying power) of an instrument.

11.65 Answer the following questions:

(a) The angle subtended at the eye by an object is equal to the angle subtended at the eye by the virtual image produced by a magnifying glass. In what sense then does a magnifying glass provide angular magnification?

(b) In viewing through a magnifying glass, one usually positions one's eyes very close to the lens. Does angular magnification change if the eye is moved back?

(c) Magnifying power of a simple microscope is inversely proportional to the focal length of the lens. What then stops us from using a convex lens of smaller and smaller focal length and achieving greater and greater magnifying power?

(d) Why must both the objective and the eyepiece of a compound microscope have short focal lengths?

(e) When viewing through a compound microscope, our eyes should be positioned not on the eyepiece but a short distance away from it for best viewing. Why? How much should be that short distance between the eye and eye-piece?

11.66 An angular magnification (magnifying power) of 30× is desired using an objective of focal length 1.25 cm and an eyepiece of focal length 5 cm. How will you set up the compound microscope?

11.67 A small telescope has an objective lens of focal length 140 cm and an eyepiece of focal length 5.0 cm. What is the magnifying power of the telescope for viewing distant objects when

(a) the telescope is in normal adjustment (i.e. when the final image is at infinity),

(b) the final image is formed at the least distance of distinct vision (25cm).

11.68 (a) For the telescope described in 11.67 (a), what is the separation between the objective lens and the eyepiece?

(b) If this telescope is used to view a 100 m tall tower 3 km away, what is the height of the image of the tower formed by the objective lens?

(c) What is the height of the final image of the tower if it is formed at 25 cm?

11.69 An amateur astronomer wishes to estimate roughly the size of the Sun

using his crude telescope consisting of an objective lens of focal length 200 cm and an eye-piece of focal length 10 cm. By adjusting the distance of the eye-piece from the objective, he obtains an image of the sun on a screen 40 cm behind the eye-piece. The diameter of the sun's image is measured to be 6.0 cm. What is his estimate of the sun's size, given that the average earth-sun distance is 1.5×10^{11}m.

11.70 (a) List some advantages of a reflecting telescope, especially for high resolution astronomy.

(b) A reflecting telescope has a large mirror for its objective with radius of curvature equal to 80 cm. What is the magnifying power of the telescope if the eye-piece used has a focal length of 1.6 cm?

11.71 (a) The image of the objective in the eye-piece is known as the 'eyering'. Why is this the best position of our eyes for viewing?

(b) Show that the angular magnification of a telescope equals the ratio of the diameter of objective to the diameter of eye-ring.

(c) The angular magnification of a telescope is 300. What should be the diameter of the objective if our eyes (located at the eyering) are just able to collect all the light refracted by the objective? Take the diameter of the pupil of the eye to be 3 mm.

11.72 A terrestrial telescope has an objective of focal length 180 cm and an eye-piece of focal length of 5 cm. The erecting lens has a focal length of 3.5 cm. What is the separation between the objective and the eye-piece? What is the magnifying power of the telescope? Can we use the telescope for viewing astronomical objects?

11.73 (a) A Galilean telescope obtains the final image erect (like in a terrestrial telescope) without an intermediate erecting lens. It does so by using a diverging lens for its eye-piece. Show that the angular magnification of a Galilean telescope is given by a formula similar to that for any ordinary telescope: angular magnification$= -f_o/f_e$ (negative sign because f_e is negative).

(b) For a Galilean telescope with f_o = 150 cm, $f_e = -7.5$ cm, what is the separation between the objective and the eye-piece?

(c) What is the main disadvantage of this type of telescope?

11.74 Describe briefly the construction of prism binoculars. Explain with the help of a ray diagram how the inversion of the image produced by an ordinary telescope is reversed by the use of two right-angle totally reflecting prisms. Explain carefully how the final image is both erect and without any 'lateral inversion'. List some advantages of prism binoculars over an ordinary telescope.

11.75 The objective of telescope A has a diameter 3 times that of the objective of telescope B.

(a) How much greater amount of

light is gathered by A compared to B?

(b) Show that the range of A is three times the range of B. (Range tells you how far away a star of some standard absolute brightness can be spotted by the telescope).

(c) A telescope increases the brightness of the background compared to what is seen by the unaided eye. Thus it facilitates observation by improving the contrast between a star and its background. Explain this statement carefully.

11.76 A 35 mm slide with a 24 mm × 36 mm picture is projected on a screen placed 12 m from the slide. The image of the slide picture on the screen measures 1.0 m × 1.5 m. Determine the location of the projection lens, and its focal length.

11.77 In order to help you understand clearly the functions of the condensing lens and the projection lens in a projector, the following questions have been devised. Answer them carefully:

(a) Why do we need a condensing lens at all? Can we not project a slide simply by illuminating it by a lamp and obtaining its magnified image on a screen?

(b) For projecting the slide in 11.76, how much should be the least diameter of the condensing lens? Where should the slide be placed relative to this lens?

(c) The image of the source formed by the condensing lens should be located on the projection lens. Why? What is the preferred size of this image of the source? Explain.

(d) The condensing lens converges light from the source on to the slide. Does that mean the image of the source is formed on the slide? If not, why not?

(e) Consider a small portion of the image on the screen, say the lowermost portion. Does this portion get illumination only from the lowermost part of the source? Or from all points of the source? Explain.

11.78 The following questions will help you clearly understand the simple optical principles involved in a camera. Answer them carefully:

(a) Explain the term 'depth of field'. Why does the depth of field increase if aperature is reduced? Which shot in your view will require a greater depth of view? Photograph of a scenic spot, your identity photograph?

(b) The filed of view of a camera is increased by using a so-called 'wide-angle lens'. In what way does this lens differ from an ordinary camera lens? How does it increase the field of view?

(c) What is a telephoto lens? How does it differ from an ordinary camera lens? How does it increase the field of view?

(d) Why are aperatures of camera lenses so small while the apera-

tures of telescopes are as large as feasible?

11.79 (a) Show that for a given brightness of the image on a camera film, the exposure time t is inversely proportional to the square of the aperature size a and directly proportional to the square of the focal length f of the camera lens.

(b) A camera is set at the operature size $f/8$ and the exposure time of (1/60)s. How much exposure time is required for receiving the same amount of light if the operature size is set at $f/5.6$? How is the depth of field affected by this change?

11.80 An eye-piece of a telescope consists of two plano-convex lenses L_1 and L_2 each of focal length f separated by a distance of $2f/3$. Where should L_1 be placed relative to the focus of the objective lens of the telescope so that the final image through L_2 is seen at infinity?

[*Note:* An arrangement of lenses like this is preferred to a simple double-convex lens for an eye-piece because it reduces chromatic and spherical aberrations. Details are beyond our scope here.]

Albert Abraham Michelson (1852-1931) Outstanding American experimenter who measured the speed of light with steadily improving accuracy over a long career, ultimately reaching within 1 kilometer per second of the presently accepted value. His experiment with Morley failed to find changes in the speed of light produced by motion with respect to the ether and thus laid the foundation for special relativity. The interference arrangement which he invented for this purpose was put to good use in spectroscopy, giving a wavelength resolution of a thousandth of a nanometer. His work with Pease on the coherence of starlight collected from two mirrors upto 20 ft. apart gave the first measurement of the angle subtended by any star (less than 10^{-6} radian!).

CHAPTER 12

Electrons and Photons

12.1 Introduction

So far, we have not enquired too closely into the source of electricity and magnetism. In this Chapter and the next, we go into this question. Attempts to answer it have led to the discovery of completely new phenomena, and equally new ideas necessary to make sense of observations. This revolution, which started with the experimental study of electrical discharges in gases, and the discovery of the electron, is broadly called modern physics. We shall present here an elementary account of electrons - a basic component of atomic matter, and of photons, the particle aspect of electromagnetic waves. In the next Chapter, we briefly describe atoms, molecules and nuclei.

Electrons play a key role in all chemical properties of atoms and molecules, and in the physical properties of condensed matter (solids and liquids). One such property is electrical conduction. Therefore, the

concept of electrons was introduced in the course in chemistry and in the Chapters on electricity and magnetism before describing the direct experimental evidence for their existence here in the physics text book.

Electronics has become a common household word. The television, the transistor radio, calculators and digital watches are all devices based on the conduction of electrons. Although no one has "seen" an electron as we see each other, its existence is well established. How did we come to know of it? Exploration in this unknown territory involved chance discovery, false clues, dogged hard work, inference from facts, forming tentative hypotheses and their test through experiments. The story of the discovery of the electron is as exciting as any of your favourite detective stories.

Experiments which led to the discovery of electrons had their source in Faraday's investigation of passing an electric current in a rarefied tube containing air. He observed a purple glow in the tube. Such a phenomenon is called an electric discharge. Faraday's phenomenon undergoes changes in form and nature if the gas contained in the tube is rarefied further. At a given low pressure of gas, below 2 Pa, rays are emitted from the negative pole (cathode) which are invisible to the naked eye but which can be discovered through certain peculiar effects. When these rays hit the walls of the glass tube, faint glow of light are seen at the point of impact. The phenomenon of emission of visible radiation by glass surfaces, especially when coated with material like ZnS, under impact of charged particles or on illumination by ultraviolet light is called fluorescence. The rays also produce sharp shadows when objects such as a mica cross are put in their paths. Therefore, like rays of ordinary light they propagate in straight lines. They differ from light as they can be deflected from their straight paths by means of a magnet. These rays were given the name *cathode rays* for obvious reasons. By the end of the nineteenth century, two views were prevalent about the nature of cathode rays: one, that the rays are negatively charged bodies or corpuscules shot off from the cathode with great velocity; the other that the rays are some kind of vibrations or waves.

Phillip E.A. von Lenard and Joseph John Thomson were awarded the Nobel prize in physics in 1905 and 1906, respectively for their work on cathode rays. Lenard in the opening remarks of his Nobel lecture gave credit to the work of his predecessors and spoke of not only the fruits but also of the trees which have borne them, and of those who planted the trees. In Table 12.1, we have listed the scientists who were associated with this orchard along with a citation of the fruits of their labours.

By the end of the nineteenth century, Thomson's identification of the cathode rays in a discharge tube as a beam of streaming electrons was generally accepted. Thomson measured the ratio of charge and mass of the electron. Millikan followed up his work and measured the individual electronic charge (not of cathode rays, though) and observed that charges occur in integral multiples of a fundamental charge which is identified with the charge of an electron.

C.J. Davisson shared the 1937 Nobel prize in physics with G.P. Thomson, the son of illustrious J.J. Thomson for the discovery of electron waves. Davission in his Nobel address expressed so beautifully the excitement of physics that took place during the first twentyfive years of the twentieth century that we reproduce a paragraph from it for your pleasure.

"Troubles, it is said, never come singly and the trials of the physicist in the early years of the century give grounds for credence in the

Table 12.1:

Faraday	(1838)	Sent a current through a glass tube containing air at low pressure and observed the purple glow extending from the positive electrode almost to the negative electrode.
Plücker	(1858)	Observed that the purple glow disappears at very low pressure with the tube becoming dark and that emissions from the cathode called cathode rays respond to a magnetic field.
Hittori	(1869)	Placed solid objects in the discharge tube and observed sharp shadows on the glass wall.
Goldstein	(1876)	Observed that the cathode rays are emitted perpendicular to the surface of any cathode and the properties of the rays are independent of the nature of the cathode material.
Crookes	(1879)	Designed a glass tube called the Crookes tube for producing cathode rays free from interferences. Crookes obtained a sharp shadow by placing a mica cross between the cathode and the walls of the tube.
Hertz and Lenard	(1887)	Measured the speed of propagation of the cathode rays and studied their propagation through air and metal foils.
Kaufmann	(1897)	Measured e/m.
Thomson	(1897-1898)	Measured e/m.
Millikan	(1906)	Measured individual electronic charge.

pessimistic saying. Not only had light, the perfect child of physics, been changed into a gnome with two heads - there was trouble also with electrons. In the open they behaved with admirable decorum, observing without protest all the rules of etiquette set down in Lorentz' manual, but in the privacy of the atom they indulged in strange and unnatural practices; they oscillated in ways which no well-behaved system would deem proper. What was to be said of particles which were ignorant apparently of even the rudiments of dynamics? Who could apologise for such perversity - rationalise the data of spectroscopy? A genius was called for, and a genius appeared. In 1913 Niels Bohr gave us his strange conception of 'stationary' orbits in which electrons rotated endlessly without radiating, of electrons disappearing from one orbit and reappearing, after brief but unexplained absences, in another. It was a weird picture - a picture to delight a surrealist - but one which fascinated the beholder, for in it were portrayed with remarkable fidelity the most salient of the orderly features which spectroscopic data were known to possess; these were the Balmer series! and there was the Rydberg constant! - correct to the last significant digit! It was a master piece. It is important to note that in achieving this tour de force Bohr made judicious use of the constant which Planck had extracted from the black body spectrum, the constant h".

Thus, by the end of the first quarter of the twentieth century, enough evidence on strange ideas such as wave-particle nature of light and particle-wave nature of matter had accumulated. These ideas conflict with common sense. The two-hundred years old Newtonian mechanics was shaken from its foundations and gave way to a new mechanics known as the quantum mechanics. In common parlance the physics described in this Chapter and the next is called modern physics.

A large number of Nobel prizes in physics were awarded during the first forty years of the twentieth century for discoveries and ideas which constitute the modern physics. At the end of this Chapter we have given a table (Table 12.3) listing some of the awards in chronological order relating to the concepts given in Chapters 12 and 13.

Although in the framework of quantum mechanics, "seeing" an electron as we see each other has no meaning, we shall study next the phenomena of conduction of electricity through gases and the experiments of J.J. Thomson which revealed the existence of electrons.

12.2 Discharge through gases at low pressures

The phenomenon of discharge through gases is a familiar sight in the big cities where one comes across advertising signs in red (neon discharge) and pale green (helium discharge). The household fluorescent light is also a discharge through a gas at low pressures.

The experimental arrangement for showing the discharge through air at low pressures is as shown in Fig. 12.1. A discharge tube made out of Pyrex glass and about 40 cm long is connected to a mechanical rotary pump. The two circular electrodes fixed at the ends of the discharge tube are connected to an induction coil which can supply a potential difference of several thousand volts across the tube. The electrode connected to the negative terminal of the induction coil is called the cathode and the electrode at the high positive potential is called the anode.

At atmospheric pressure in the tube there is no visible effect. The medium in the tube is then non-conducting. As soon as the vacuum pump is switched on, a pink glow ex-

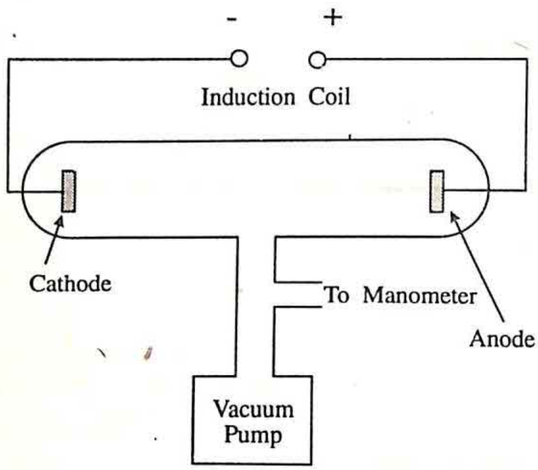


Figure 12.1: Experimental arrangement for showing the discharge through air at low pressure.

tending from the anode almost to the cathode is seen. This glow is the forefather of the modern neon light. As the tube is further evacuated, many striking visible effects such as striations in the discharge column are seen. When pressure in the tube is lowered below 2 Pa the interior of the tube becomes dark. The tube is said to black out at this stage. The charged particles constituting the cathode rays excite atoms of the gas in the tube by collision processes. The excited atoms emit light with colours characteristic of their structure. As the pressure of the gas is lowered the density of the gas decreases and cathode rays are able to travel from one end of the tube to another without meeting any atoms, i.e. without undergoing appreciable collisions with the atoms. The atoms of the gas remain unexcited and therefore the discharge tube at this stage blacks out. By his careful experiments, Thomson convincingly resolved the controversy about the nature of the cathode rays which we describe next.

12.3 Cathode rays

Experiments which have led to the discovery of electrons involve the blackout stage of the discharge tube. Thomson, who had devoted much of his scientific life to studying the phenomenon of electrical conduction through gases at low pressures, used discharge tubes of various shapes. For measuring the properties of the cathode rays, Thomson used a discharge tube which resembles the modern TV tube.

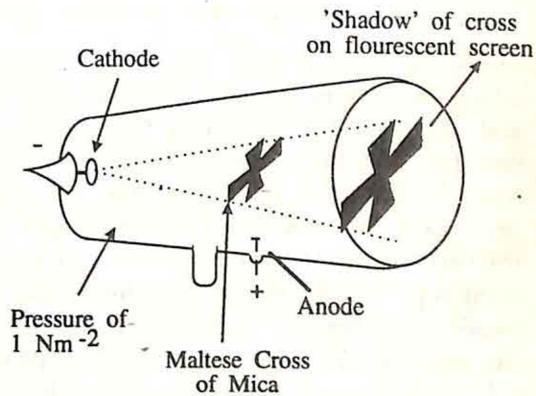


Figure 12.2: Crookes tube with anode at a side for demonstration of rectilinear propagation of cathode rays.

In the Crookes tube used by Thomson, Fig. 12.2, the anode is on one side so that it does not obstruct the emissions from the cathode which reach the fluorescent screen. Emissions from the cathode which cause fluorescence were called the cathode rays. By inserting metallic objects of different geometrical shapes and observing their sharp shadows on the fluorescent screen, Thomson concluded that cathode rays travel in straight lines. These observations are consistent with the hypothesis that cathode rays

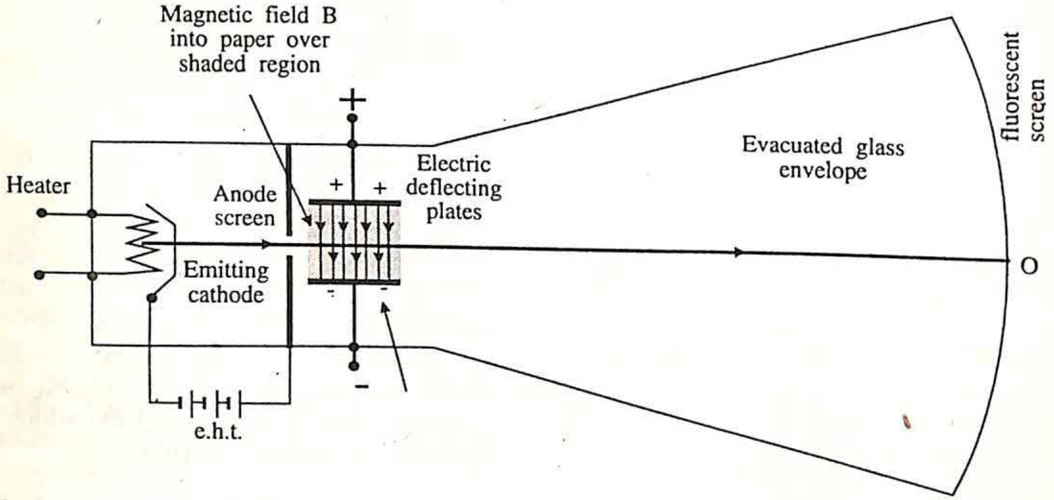


Figure 12.3: Thomson's apparatus for the measurement of e/m.

are some type of short wavelength waves which can travel through evacuated media and also with the possibility that they are particles. Thomson also made a rough estimate of the speed of cathode rays. For this he marked two scratches along the length of the cathode ray tube and measured the time delay between the fluorescence glows using a rapidly rotating mirror. He estimated that the rays travelled with a velocity of 2×10^5m s^{-1}, which is much smaller than the velocity of electromagnetic radiation.

By applying electric and magnetic fields across the discharge tube, Thomson observed that cathode rays are deflected as though they consist of moving negative electric charges. You know that charges in motion experience acceleration in a magnetic field if the field is not parallel to the direction of the motion. The direction and magnitude of the force on a moving charge is given by the Lorentz force formula. On the other hand, a ray of light is not bent by external electric and magnetic fields. Goldstein had observed that these rays are emitted perpendicular to the surface of the cathode and are independent of the material of which it is made. If the cathode rays were electromagnetic radiation they would be emitted in all directions from the cathode. These observations clearly rule out the initial opinion that these cathode rays are electromagnetic waves of short wavelength. Thus, the hypothesis that cathode rays are moving negative electric charges, called electrons, appears very plausible and has been accepted as the proof of the existence of electrons and their role as the building blocks of atoms.

Thomson modified the design of his discharge tube as shown in Fig. 12.3. Now both electric and magnetic fields could be applied perpendicular to the direction of propagation of the cathode rays and to each other. By this ingenious method, Thomson made a crude estimate of the charge to mass ratio, e/m of the electrons. We describe next the basic physics of his experiment.

12.4 e/m of Electrons

Let the accelerating voltage between the cathode and the anode be V volts. Let the direction of propagation of the electrons in the cathode ray tube be called the x-axis. On reaching the anode, the electrons escape

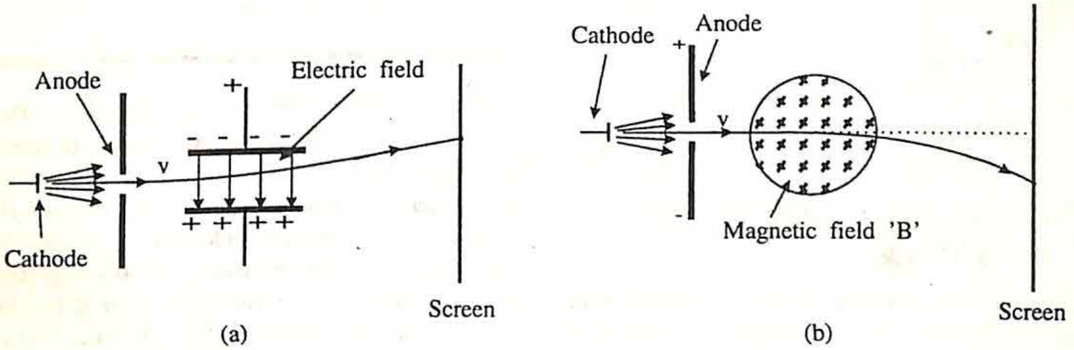


Figure 12.4: (a) Deflection of the spot upward when electric field is applied along the negative y-axis. (b) Deflection of the spot downward when magnetic field is applied in the negative z-direction.

through a hole and are collimated as a narrow beam. On reaching the screen the beam marks a fluorescent spot.

If a uniform electric field is applied along the negative y-direction (downward with respect to the screen) the spot on the screen is seen to get deflected upward as shown in Fig. 12.4a. This is easy to understand because the force on a particle of charge e in an uniform electric field $\mathbf{E}$ is $e\mathbf{E}$. Since the electrons are negatively charged, cathode rays are deflected in directions opposite to the $\mathbf{E}$ field. If a uniform magnetic field $\mathbf{B}$ is applied in the negative z-direction, the spot on the screen is seen to deflect downward on the screen, see Fig. 12.4b. You may recall that force on a moving charge in an external magnetic field is perpendicular to both the direction of velocity vector and the direction of the field. The magnitude of the force is evB. If the strengths of the $\mathbf{E}$ and $\mathbf{B}$ fields are adjusted such that the cathode ray beam falls on the screen at its undeflected spot the actions of electric and magnetic fields completely cancel each other (Fig. 12.3). The mathematical expression of this condition is

$$eE = evB, \tag{12.1}$$

giving

$$v = \frac{E}{B}. \tag{12.2}$$

In this expression neither the charge of the electron e, nor its mass, m appear. What extra information did Thomson use for obtaining the ratio e/m?. He used the speed of electrons as they emerge out of the anode. This can be estimated by using the principle of conservation of energy.

The electrons at the cathode possess potential energy of eV joules higher than at the anode. V is the potential difference between anode and cathode. As the electrons travel toward the anode under the action of the electric field in the space between the cathode and the anode their potential energy is converted into kinetic energy. The situation is similar to dropping a ball in the gravitational field of the earth. Therefore, the speed of the electrons at the anode can be estimated from the equality,

$$\frac{1}{2}mv^2 = eV, \tag{12.3}$$

giving

$$v = \sqrt{\frac{2eV}{m}} \tag{12.4}$$

Combining Eqs (12.2), (12.3) and (12.4), we

get

$$\frac{2eV}{m} = \frac{E^2}{B^2}$$

or

$$\frac{e}{m} = \frac{E^2}{2VB^2}. \qquad (12.5)$$

Thomson's first results for e/m was

$$0.77 \times 10^{11}\text{Ckg}^{-1}.$$

This is considerably in error. The present accepted value of e/m, measured with more advanced apparatus is

$$\frac{e}{m} = 1.759 \times 10^{11}\text{Ckg}^{-1}. \qquad (12.6)$$

The electric charge of electrons is negative as observed from the acceleration of the cathode rays in the external electric and magnetic fields.

Example 12.1: **Electrons which have been accelerated by a potential difference of 1000 V enter electric and magnetic fields of strengths 20 V/cm and 1 G respectively as in the Thomson's experiment. Estimate the e/m and the speed of the electrons if they move undeviated through this system.**

Answer: Electric field $E = 20$ V/cm

$$= \frac{20}{10^{-2}} = 2 \times 10^3\,\text{V/m}$$

$$B = 1\text{G} = 10^{-4}\text{T}.$$

Potential difference $V = 10^3$ V.

Therefore,

$$\frac{e}{m} = \frac{E^2}{2VB^2} = \frac{(2 \times 10^3)^2}{2 \times 10^3 \times 10^{-8}}\,\text{C/kg}$$
$$= 2 \times 10^{11}\text{C/kg}.$$

As the electrons travel undeflected the speed v of the electrons can be obtained from the relation,

$$v = \frac{E}{B}$$

It gives

$$v = \frac{2 \times 10^3}{10^{-4}}\text{m/s} = 2 \times 10^7\text{m/s}.$$

You can convince yourself that by using Thomson's method it is not possible to measure the absolute value of the charge of the electrons. In fact, by measuring the acceleration under external electric or magnetic fields one can only measure e/m because the Lorentz force is directly proportional to the electric charge. According to Newton's second law, equation of motion of a charge e in an electric field E is

$$ma = eE, \qquad (12.7a)$$

and the equation of motion of a particle of charge e moving with speed v in a transverse magnetic field B is

$$ma = evB. \qquad (12.7b)$$

You can see that in *both* expressions, Eqs (12.7a) and (12.7b) the acceleration 'a' occurs in the combination e/m.

The American physicist R. A. Millikan (1868-1953) used an imaginative method to measure the charge of the electrons. He was awarded the Nobel prize in 1923 for experiments on the measurement of electronic charge. To this we turn next.

12.5 Millikan's method for measuring the fundamental charge

Although the SI unit of electric charge is the coulomb, the electrolysis experiment of Faraday (Chapter 4) had revealed that the total charge of one Avogadro number, 6.02×10^{23}, of ions is 96,500 C. From this fact, we estimated the existence of a fundamental charge e. The magnitude of e is $96,500/(6.02 \times 10^{23}) \simeq 1.6 \times 10^{-19}$C. This result can be generalized as a universal property of the electric charge called the quantization of electric charge. The quantization of electric charge

was explained in the Chapter 1. It simply means that total charge on any body can only be an integral multiple of e, i.e. $\pm e, \pm 2e, \pm 3e., \cdots$. Millikan who is regarded as one of the outstanding experimentalists of all times set up an ingenious experiment for testing the hypothesis of the quantization of charge and for measuring the value of the fundamental charge.

In the following we have described Millikan's classic experiment on the measurement of fundamental charge and reproduced and analysed the data reported by Millikan in his Nobel prize lecture. The key idea of his experiment was the measurement of the motion of charged oil drops under free fall and in a uniform electric field.

In his fundamental experiments, Millikan had arranged two parallel metal plates C and D with a small separation like in a parallel plate capacitor as shown in Fig. 12.5. The plates of the capacitor could be connected to a source of high tension (high voltage) or could be short circuited using a double switch. There was a minute pin-hole in the middle of the top plate. Millikan sprayed oil-droplets of radius about one thousandth of a millimeter. Sooner or later an oil droplet which had been charged by friction while blowing the spray would fall through the pin-hole and enter the space between the plates. The drop could be seen through a telescope like a bright star on a black background. In the eye-piece of the telescope Millikan had provided three cross-hairs at known separations. He measured the time taken by the droplet to pass between them. In this way he measured the terminal velocity of free fall, which for small droplets is only a fraction of a millimeter per second. By charging the capacitor plates to a high potential difference the droplets were seen to be pulled up by the attraction of the upper plate.

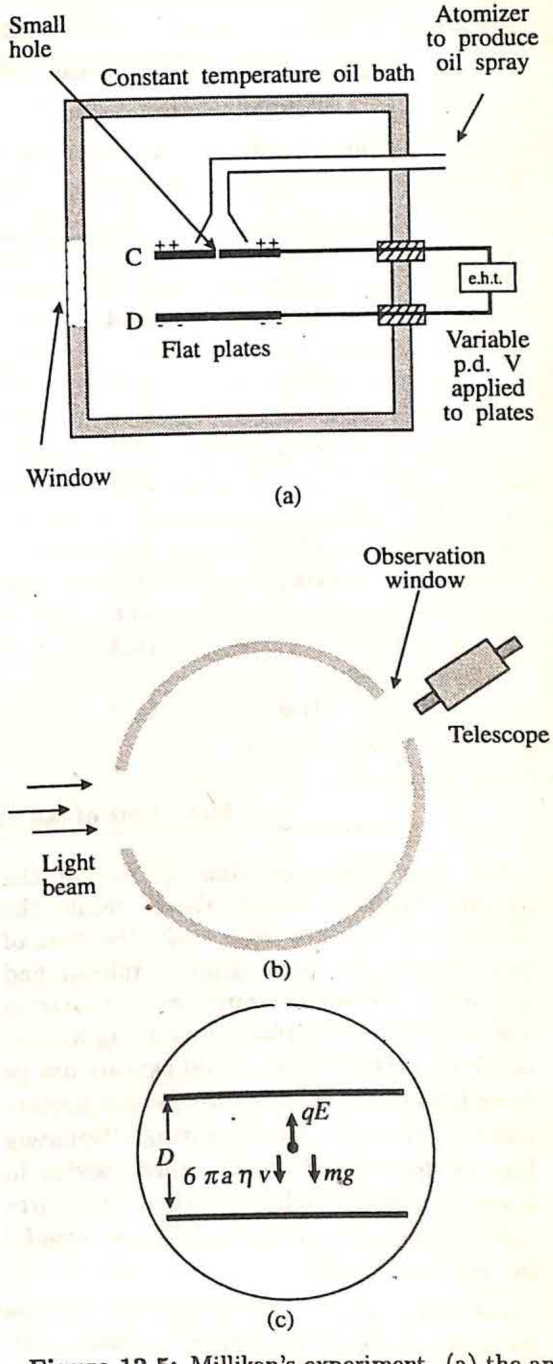


Figure 12.5: Millikan's experiment. (a) the apparatus, (b) plan view and (c) the forces acting on an electron.

Table 12.2:

Time of fall 1.303 cm under gravity (s) t_g	Time of rise 1.303 cm in field (s) t_e	Mean time of rise in field (s) t_e	$\frac{1}{t_g}+\frac{1}{t_e}$ (s^{-1})	Divisors of
120.2	68.0			
119.9	67.8	67.7	23.07×10^{-3}	1
119.8	67.4			
120.8	26.2			
120.2	26.4	26.4	46.19×10^{-3}	2
120.1	26.6			
-	26.4			
121.2	16.5			
120.1	16.3	16.55	68.99×10^{-3}	3
-	16.6			
-	16.4			
121.0	11.9	11.9	92.34×10^{-3}	4
Mean time of fall under gravity = 120.35s				

After measuring the time of rise of the droplet Millikan would short circuit the plates and measure once again the time of free fall for the same drop. Millikan had also made a special arrangement for ionising the air between the plates by shining X-rays. Sometimes the droplet would capture one or more ions from the air molecules and acquire different charges. By measuring alternately the velocity of the droplet when moving in a known electric field and when under free fall, Millikan observed the *same oil droplet for several hours*!

Let the radius of the droplet be a, m be its mass and q be the charge carried by it. Let the viscosity of air be η. According to Stoke's law (Class XI book, Section 10.6, Eq. (10.20)) the terminal velocity under free fall, v_g, and the terminal velocity, v_e, in a uniform electric field E, will be given by

$$6\pi\eta a v_g = mg \quad (12.8a)$$

$$qE = 6\pi\eta a v_e + mg \quad (12.8b)$$

(Remember that in an electric field, the drop moves up, so the viscous force is downward). Solving Eqs (12.8a), and (12.8b), you can easily obtain

$$q = \frac{mg}{Ev_g}(v_e + v_g). \quad (12.9)$$

Let D be the separation between two cross-wires and t_g, t_e be the time taken by the droplet to cover D when under free fall and when moving in the electric field E. The velocities v_g and v_e will be (D/t_g) and (D/t_e),

respectively. In terms of t_g and t_e, Eq. (12.9) can be reexpressed as

$$q = \frac{mg}{E} t_g \left(\frac{1}{t_e} + \frac{1}{t_g} \right) \qquad (12.10)$$

Suppose we assume that in nature charge occurs only in multiples of a fundamental charge e then $q = ne$, when n is some integer. Similarly when the same droplet has acquired charge q', then q' will be equal to $n'e$, where n' is some other integer.[1] It means that ratio of different charges carried by the same oil drop must be expressible as ratio of integers. This condition can be expressed as follows:

$$\frac{q'}{q} = \frac{(1/t_e') + (1/t_g)}{(1/t_e) + (1/t_g)} = \frac{n'}{n}. \qquad (12.11)$$

In Table 12.2, we have reproduced data on t_g and t_e for a given droplet from Millikan's Nobel prize lecture (page 57, *Nobel Lectures Physics* 1922-1941, Elsevier Publishing Company, 1965).

While calculating mean time of rise in field, Millikan had averaged similar data such as 16.5, 16.3, 16.6, 16.6 as they correspond to situations when the oil droplet had the same charge. You can easily note from the last column that the data agrees with the hypothesis that charge can occur only as integral multiples of a fundamental unit. In the data given above the droplet had charges of $e, 2e, 3e$, and $4e$ units.

The absolute value of the fundamental charge can be obtained from Eq. (12.10) by substituting the value of the electric field E. From the analysis of his data, Millikan found the value of electron charge e to be 1.591 $\times 10^{-19}$C. The fundamental charge has now been measured to an accuracy of eight decimal places. Its value is

$$e = 1.6021892 \times 10^{-19}\text{C}.$$

Is the magnitude of Millikan's fundamental charge e same as the charge carried by Thomson's electrons? Enough evidence has piled up from a whole web of interconnected experiments and theories, which makes sense only if the charge in Thomson's e/m and Millikan's fundamental charge are the same. You will have to wait till you study in the next Chapter Niels Bohr's calculation of the Rydberg constant by substituting this value of the electron charge and values of other fundamental constants in his celebrated formula and its reliable agreement with the value obtained by spectroscopic analysis of the wavelengths of light from hydrogen atoms. Therefore, if we assume that the electrons which are the consituents of cathode rays carry the fundamental charge, then from the Thomson's measured value of e/m we can estimate the mass of the electron. It gives,

$$m = \frac{1.602 \times 10^{-19}\text{C}}{1.759 \times 10^{11}\text{C/kg}} = 9.1 \times 10^{-31} \text{ kg}$$

Example 12.2: In one of his experiments, Millikan found that an oil drop took 120.35s to fall under gravity a distance of 1.303cm and 67.73s to rise through the same distance when an electric field of 6000 V/cm was applied to the plates separated by 1.6 cm. Calculate the radius and mass of the oil drop. How many electronic charges were there in the oil drop?
(oil density = 0.9199 g/cm^3, air viscosity = 182.4 $\times 10^{-6}$ g/cm s; acceleration due to gravity = 980.3 cm/s^2.)

[1] The mass of the oil drop is greater than that of ions by several orders of magnitude. As the oil drop acquires different charges by capturing or releasing ions, the change in its mass when it has different charges can be approximated to be neglegible. The time of free fall, t_g, of the same oil drop will, therefore, remain the same throughout the experiment.

Answer: Substituting $m = \frac{4}{3}\pi a^3 \rho$ in the Eq. (12.8a) and solving we get

$$a^2 = \frac{9}{2}\left(\frac{\eta v_g}{\rho g}\right), \text{ where}$$

$$\rho = \text{oil density} = \frac{0.9199 \times 10^{-3}\text{kg}}{10^{-6}\text{m}^3}$$

$$= 0.9199 \times 10^3 \text{ kg/m}^3$$

$$\eta = 182.4 \times 10^{-6}\text{g/cm s}$$

$$= \frac{182.4 \times 10^{-6} \times 10^{-3}\text{kg}}{10^{-2}\text{m s}}$$

$$= 182.4 \times 10^{-7}\text{kg/m s}$$

$$a^2 = \frac{9 \times 182.4 \times 10^{-7}}{2 \times 0.9199 \times 10^{-3}} \times \frac{1.303 \times 10^{-2}/120.35}{9.803}$$

$$a = 9.927 \times 10^{-7}\text{m.}$$

$$m = \frac{4}{3}\pi a^3 \rho = \frac{4}{3}\pi (9.927 \times 10^{-7})^3 \times 0.9199 \times 10^3$$

$$= 3.7695 \times 10^{-15}\text{kg}$$

From Eq. (12.10); $q = (mg/E)(1 + (t_g/t_e))$

$$q = \frac{3.7695 \times 10^{-15} \times 9.803}{(6000/10^{-2})} \times \left(1 + \frac{120.35}{67.73}\right)$$

$$= 1.70 \times 10^{-19}\text{C}$$

Electronic charge $e = 1.6 \times 10^{-19}$C

Number of electronic charges on the drop

$$= \frac{q}{e} = \frac{1.70 \times 10^{-19}\text{C}}{1.60 \times 10^{-19}\text{C}} \simeq 1.$$

The oil drop carried 1 electronic charge.

Comparing the mass of the electron with the mass of ionised hydrogen atom (proton) we see that it is lighter by a factor of 1836. This indicates that the electrons are easier to accelerate than ions. The discovery of the electron as representing a free electric charge and the possibility of its extraction or absorption by a neutral atom to form ions provided the first experimental evidence that atoms have structure and consist of positive and negative charges.

12.6 Free electrons in metals

It is now well established that all electrons are identical and are one of the fundamental constituents of atoms. In Chapter 10 of the Class XI textbook on the Section on solids, we have introduced the concept of crystalline structures. In crystalline solids atoms are arranged in three dimensional periodic patterns. In some solids called metals, there are electrons in each atom so loosely attached to their parent that they can move from one atom to another atom and in fact hop around throughout the solid. Such electrons are called free electrons. In Chapter 3 on current electricity, we have pointed out that electrical current in metals is carried by these free electrons. There, we have also explained why these free electrons though free to move inside the metal do not escape outside the metal surface. These electrons are held inside the metal surface by the attraction of the ions on the lattice sites and because of surface forces, and therefore require a minimum amount of energy in order to spill out of the surface. This energy is called the *work function* of the metal. This minimum energy can be supplied to the free electrons in the metal for their release from the metal surface by any one of the following physical processes:

(a) *Thermionic emission*: - by heating the metal sufficient thermal energy can be given to free electrons to overcome the attractive pull of the metal surface.

(b) *Field emission*: - electrons can be extracted from metals by applying an electric field.

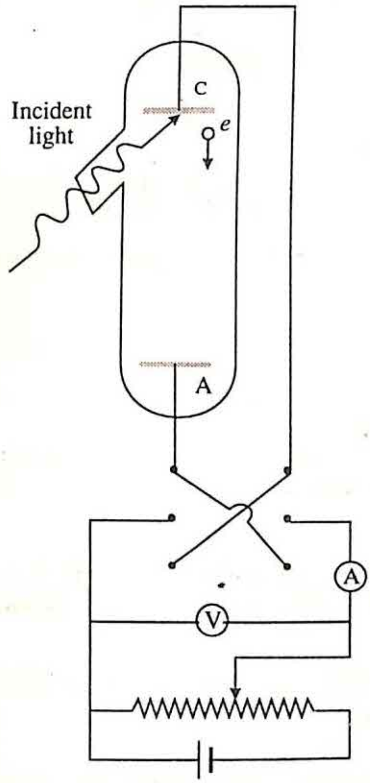


Figure 12.6: Experimental arrangement for observing photoelectric effect.

(c) *Photoelectric emission*: - by shining light of high frequency (ultraviolet) on clean metal surfaces, electrons from inside the metal can be released.

We shall next study the photoelectric effect. Einstein explained it on the basis of Max Planck's quantum idea. This laid the foundation of the quantum theory. Therefore, the photoelectric effect is of special interest.

12.7 Photoelectric effect

The phenomenon of photoelectric effect is studied by using an experimental arrangement shown in Fig. 12.6. When monochromatic light is incident on the cathode, electrons are emitted. If some of the electrons reach the anode A, a current flows in the external circuit. The number of emitted electron that reach the anode can be increased or decreased by making the anode positive or negative with respect to the cathode. The minimum negative (retarding) potential required to stop the electron to reach A is called *cut-off voltage*.

This effect in itself is not so remarkable. After all, free electrons in the metallic cathode can absorb energy from the electromagnetic waves impinging on them. After sufficient energy has been absorbed free electrons inside the metal should be able to overcome the combined potential barrier offered by the metal surface and the retarding potential across the phototube. When the photocurrent is measured by varying (a) the intensity of light, (b) its frequency and (c) the retarding potential between the anode and the cathode, effects are observed which cannot be reconciled with the classical wave properties of light and its absorption by electrons.

The maximum kinetic energy with which the electrons leave the cathode can be measured by adjusting the retarding potential till the photocurrent in the external circuit is reduced to zero. Then electrons are not able to reach the anode. If V is the cut-off voltage, the maximum kinetic energy of electrons in the phototube is eV.

When a careful study is made of photoemission by varying the above mentioned parameters in the experiment, the following important conclusions are reached:

(i) The energy distribution of the emitted electrons is independent of the intensity of light. That is, more photoelectrons are emitted if the intensity of light is increased but the maximum kinetic energy with which the electrons leave the metal remains unchanged. In

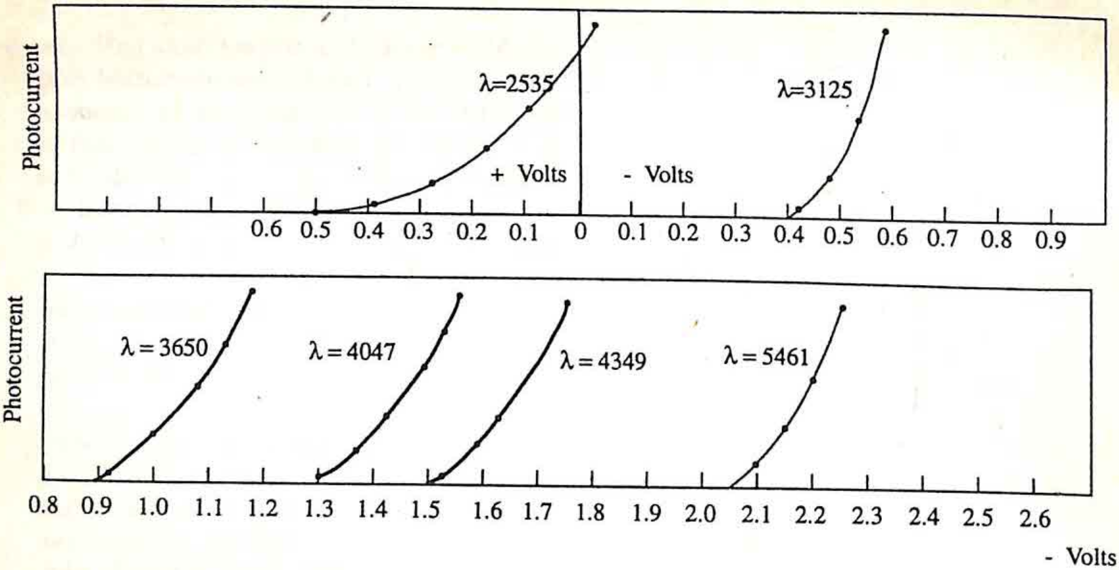


Figure 12.7: Graph of photocurrent versus voltage for different wavelengths of incident light. (Note that there was a systematic error of 2.53 V in the voltage measurement. Corrected values of the cut-off voltage have been used in Fig. 12.8).

fact, even with light of very low intensity some electrons with the same maximum kinetic energy are emitted.

(ii) Within the limits of experimental accuracy it is observed that there is no time lag between the arrival of light at the metal and the emission of photoelectrons. The delay times has been experimentally measured. It is less than 10^{-9} s.

(iii) For a given metal, photoelectrons are not emitted if the incident light is of frequency less than a critical value, called the threshold frequency, no matter how high its intensity.

(iv) The maximum kinetic energy with which photoelectrons are emitted from a particular metal and the frequency of the incident light are related linearly. The relation can be expressed as

$$KE_{\max} = h(\nu - \nu_0) \qquad (12.12)$$

where h is a constant of proportionality.

As the kinetic energy of electrons cannot be negative, photoemission does not take place when the frequency of the incident light is less than ν_0. Although the threshold frequency ν_0 changes from metal to metal, the slope of the straight line

$$eV = h(\nu - \nu_0),$$

where V is the magnitude of the cut off voltage, is the same.

Millikan also has the credit of making the first accurate measurement of cut-off voltage for sodium metal by using monochromatic light of known frequencies. We have reproduced from Millikan's published work (Nobel lecture) the graph of photocurrent versus voltage (Fig. 12.7) and the graph of cut-off voltage verses frequency of light, (Fig. 12.8). From Fig. 12.8 we can estimate the slope of

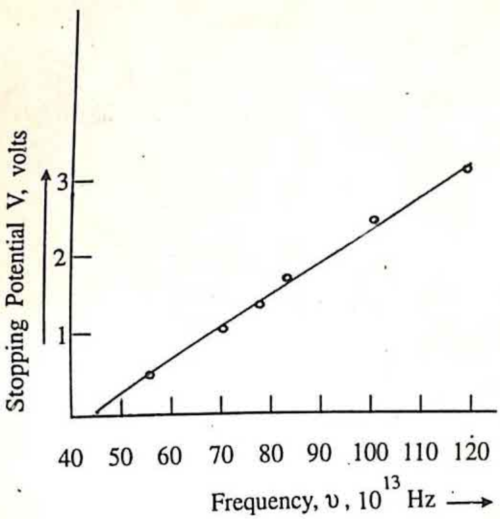


Figure 12.8: Graph of cut-off voltage versus frequency of light. The figure shows actual data for Na as reported by R. A. Millikan in 1916.

the straight line. It is

$$\frac{\Delta V}{\Delta \nu} = \frac{3\text{V}}{(121.00 - 48.25) \times 10^{13}\text{s}^{-1}}$$
$$= 4.124 \times 10^{-15}\text{V s}.$$

By multiplying it with the charge of an electron, which is the fundamental char $e = 1.602 \times 10^{-19}$C, we get

$$h = 4.124 \times 10^{-15} \times 1.602 \times 10^{-19}\text{J s}$$
$$= 6.6 \times 10^{-34}\text{J s}$$

It is numerically the same as the Planck constant h of the black body radiation formula. Planck had first introduced this constant in the celebrated empirical formula called the Planck's law for explaining the distribution of radiation of different wavelengths by a blackbody at a given temperature. This distribution was introduced to you in the Section on radiation of the Chapter 11 of your text book of Class XI (while the distribution is shown there, no formula is given for it; the formula does involve the Planck's constant h). Einstein used Planck's idea of the quantum of energy to explain all the above features of the photoelectric effect. This is discussed in the next Section.

12.8 The photon and the quantum interpretation of the photoelectric effect

Einstein took Planck's idea of the quantum of energy seriously and proposed that a monochromatic electromagnetic wave of frequency ν consists of discrete quanta, each having energy

$$E = h\nu. \tag{12.13}$$

where h is the Planck constant. The quanta of light were appropriately called photons. Each photon travels with the velocity of light. According to Einstein's special theory of relativity, which you will learn later, the energy E, and momentum, p, of particles moving with the speed of light are related:

$$E = pc, \tag{12.14}$$

where c is the speed of light.
Comparing Eqs (12.13), and (12.14), the momentum of the photon is seen to be related to the wavelength of the light as

$$p = \frac{h}{\lambda}. \tag{12.15}$$

Einstein suggested that absorption of energy from a photon by a free electron inside the metal is a single event and involves transfer of energy in one lump instead of the continuous absorption of energy as in the wave model of light. Energy is conserved in the process. It can be expressed by the relation,

Energy of the incident photon =
maximum kinetic energy of the electron
+work function of the metal. (12.16)

The kinetic energy of the emitted electron will be maximum if the free electron, which

is released from the atom, belongs to the group which has the maximum energy inside the metal. By using the Einstein relation for the energy of photons of frequency ν, we can write the photoelectric emission equation, Eq. (12.16) as

$$h\nu = KE_{\max} + \text{work function.} \qquad (12.17)$$

Let the work function be expressed in units of frequency such that

$$\text{Work function} = h\nu_0 \qquad (12.18)$$

Then the Einstein photoelectric equation, Eq. (12.17), can be re-expressed as

$$KE_{\max} = h(\nu - \nu_0).$$

This equation is identical to the experimentally observed relationship given by Eq. (12.12). Einstein received the Nobel Prize in physics in the year 1921 for the quantum theory of the photoelectric effect and not for the special theory of relativity!

The particle-like behaviour of light was also seen in elastic scattering of photons with electrons. Arthur Holly Compton investigated the scattering of monochromatic X-rays from electrons. He observed that the scattered X-rays had longer wavelength. The change in wavelength was found to be independent of the matter used for scattering but varies with the angle between the incident and the scattered rays. Compton could explain the effect observed by him by assigning momentum of magnitude $h\nu/c$ to photons of energy $h\nu$. The elastic scattering of a photon from an electron at rest can be worked out by invoking the principles of conservation of energy and conservation of momentum. The formula giving the change of wavelength of the X-ray photon is

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\varphi),$$

where φ is the angle of scattering of the X-ray photon and m is the mass of electron.

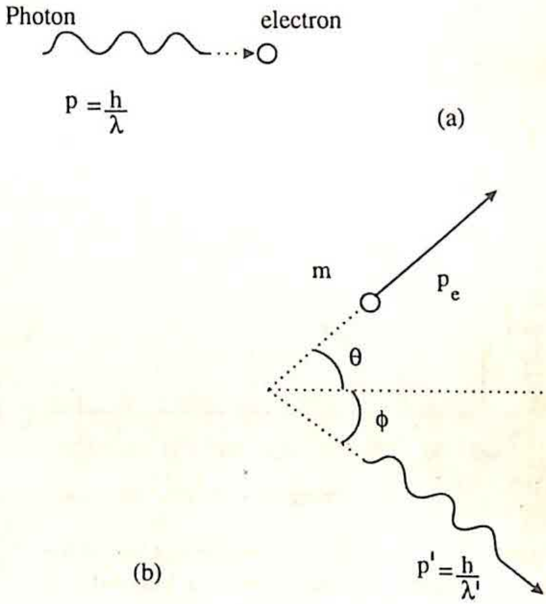


Figure 12.9: Compton scattering diagram. The figure shows (a) a photon of wavelength λ before colliding with an electron at rest and (b) the photon after collision (with wavelength λ') and the electron moving off.

The elastic scattering process is shown diagrammatically in Fig. 12.9. The recoil electrons were observed in Wilson's cloud chamber. Wilson shared the 1927 Nobel prize in physics with Compton.

You may have also appreciated the simplicity and elegance of Einstein's explanation of the photoelectric effect. He introduced revolutionary ideas which were contrary to the scientific opinion of the time. The photon hypothesis disturbed the scientific community much more than the seventeenth century Newton-Huygens heated debate on the corpuscular and the wave nature of light. But the new theory gave a better description of the physical nature than the comfortable old classical ideas.

We describe next the photo-cell which is a technological application of the photoelectric

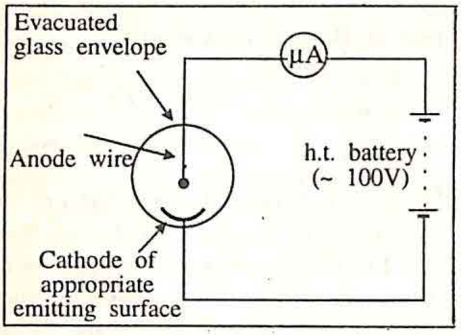


Figure 12.10: A photo-cell in a circuit.

effect.

12.9 The photo-cell

A photoemission cell in a circuit is shown in Fig. 12.10. It makes use of photo-emission from a metal surface for measureing the intensity of light. The photo electrons emitted from the cathode of the photo-cell are drawn to the collector (anode) by an electric field. The resultant electric current is measured by a sensitive meter in the external circuit. The current obtainable from a typical photo-cell is of the order of a microampere.

The fundemental use of a photo-cell is to convert a change in the intensity of illumination into a change in electric current. This change in electric current may be used to operate controls and in light measuring devices. For example, a person approaching a doorway may interrupt a light beam which is incident upon a photo-cell. The abrupt change in photocurrent may be used to start a motor which opens the door or rings an alarm. In scientific work photo-cells are used whenever it is necessary to measure the intensity of light. Light meters in cameras work on this principle.

12.10 Wave nature of matter

The discovery in 1905 of the dual nature of light was followed after two decades by the discovery of the dual nature of matter. The dual nature of light means that a monochromatic beam of light of frequency ν, hence possessing wave attributes, manifests in some experiments as though it is a stream of quanta (of energy and momentum) called photons. The energy and momentum of each photon are $h\nu$ and $h\nu/c$, respectively. The dual nature of matter is the reciprocal manifestation of a wave-like behaviour in some experiments of a stream of particles having definite energy and momentum.

In 1924 Louis de Broglie postulated that matter will also exhibit a wave nature. There was no experimental evidence for it when he made this bold hypothesis. With this hypothesis de Broglie brought to life the duality of matter on par with the duality of light. He proposed that a beam of material particles each travelling with momentum p will manifest wave character with wavelength

$$\lambda = \frac{h}{p}, \tag{12.19}$$

where h is the Planck constant. The formula is identical in form to Einstein's relation, Eq. (12.15), which gives the magnitude of momentum of photons in terms of the wavelength λ. de Broglie's hypothesis attributed a wave-like character to matter. According to this hypothesis a beam of particles each travelling with momentum p can show properties of a monochromatic wave, the wavelength λ being given by de Broglie's relation. Before we describe experiment which revealed the wave property of matter in agreement with de Broglie's hypothesis, let us first estimate the order of magnitude of the de Broglie wavelength.

The de Broglie wavelength of a 1 kg object

moving with speed of 1 m/s is

$$\frac{6.6 \times 10^{-34}\text{Js}}{1\ \text{kgms}^{-1}} = 6.6 \times 10^{-34}\,\text{m}.$$

It is indeed very small.

The acid test for the wave character is the demonstration of interference and diffraction experiments. You know very well by now that for observing interference effects the path difference in waves should be of the order of magnitude of wavelength and a diffraction effect can be seen only if the size of obstacles is of the order of the wavelength. It will be easier to pass a camel through the eye of a needle than to arrange an experiement which manifests the wave character of macroscopic objects. Therefore, let us explore the possibility of observing the wave nature of matter at the microscopic scale.

In 1927 Davisson and Germer carried out experiments to observe the wave nature of matter by using a beam of electrons, each having energy of 54 eV. What is the de Broglie wavelength of this beam?

The kinetic energy in terms of momentum p for a particle of mass m is given by the standard relation

$$KE = T = \frac{p^2}{2m}.$$

It gives $p = \sqrt{2mT}$.
In the data given

$$\begin{aligned} T &= 54\text{eV} \\ &= 54 \times 1.6 \times 10^{-19}\text{J} \\ &= 86.4 \times 10^{-19}\text{J} \end{aligned}$$

The electron mass m is 9.1×10^{-31}kg. The momentum p of electrons will be

$$\begin{aligned} p &= \sqrt{2 \times 9.1 \times 10^{-31} \times 86.4 \times 10^{-19}} \\ &= 39.6 \times 10^{-25}\ \text{kg m/s} \end{aligned}$$

and their de Bròglie wavelength,

$$\begin{aligned} \lambda &= \frac{6.6 \times 10^{-34}}{39.6 \times 10^{-25}} = 1.66 \times 10^{-10}\ \text{m} \\ &= 1.66\text{Å}. \end{aligned} \tag{12.20}$$

Diffraction of X-rays of wavelength of this order was seen by Bragg in 1913 from NaCl crystals. The diffraction effect is possible because the nearest distance between atoms in NaCl crystal is 2.8Å and the crystal lattice acts like a grating. The crystalline structure of NaCl has been discussed in Section 10.1 of the Class XI textbook.

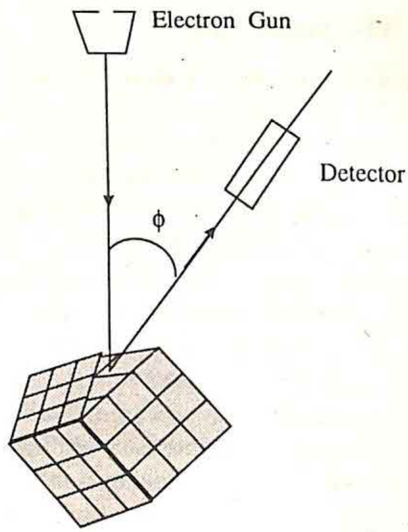


Figure 12.11: Schematic experimental arrangement used by Davisson and Germer for observing the wave nature of matter.

Davisson and Germer studied electron diffraction from a nickel crystal. The experimental arrangement has been shown in Fig. 12.11. They directed an electron beam against a (111) face of a crystal of nickel. The smallest separation between nickel atoms, d, in its crystalline structure is 0.91 Å. The intensity of the scattered electrons was measured as a function of the latitude angle ϕ measured from the axis of the

incident beam for different electron speeds. Different electron speeds corresponds to different de Broglie wavelengths. The intensity of electrons as a function of latitude angle has been reproduced in Fig. 12.12 from the Nobel lecture of Davisson. The graph shows that on scattering against the Ni surface, the 54 eV electron beam has a diffraction peak at an angle of 50° with respect to the incident direction. However, the angle of incidence and scattering relative to the Bragg planes for the case of 54 eV beam as shown in Fig. 12.12 would both be $65°(= (180° - 50°)/2)$.

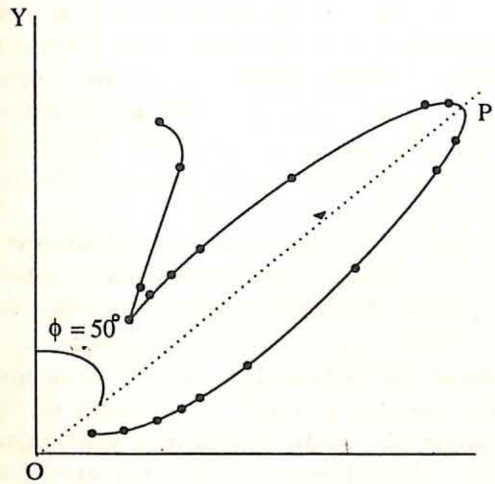


Figure 12.12: Graph showing variation in intensity of scattered electron beam as a function of latitude angle. The intensity is proportional to the length of the vector joining the origin 0 to a point (say, P) on the curve. The angle of scattering is the angle between OY and OP. The data are for electrons of energy 54 eV.

The theory of diffraction of X-rays from a crystal was worked out by Bragg. According to Bragg, X-rays of wavelength λ on scattering from a crystal of lattice spacing d show a diffraction maximum at an angle θ given by the relation

$$2d\sin\theta = \lambda \tag{12.21}$$

If the beam of electron of momentum p behaves like a wave of wavelength $\lambda(= h/p)$ its diffraction must also obey the Bragg law. We have discussed this kind of diffraction of waves from atomic planes earlier (see Section 10.5, Example 10.5). Davisson and Germer used the Bragg formula for describing the polar variation of the intensity of the diffracted electrons.

For the data of the experiment,

$$d = 0.91 \text{ Å}$$

$$\theta = 65° = 1.139 \text{ rad}$$

The wavelength $\lambda = 2 \times 0.91 \times 0.906$ Å$=$ 1.65 Å.

The measured wavelength is in close agreement with the estimated de Broglie wavelength, (see Eq. (12.20)). C. J. Davisson and G. P. Thomson shared the 1937 Nobel prize in physics for their experimental discovery of the diffraction of electrons by crystals.

This experiment convincingly established the wave nature of matter. The possibility of exploiting the wave nature of matter for designing a microscope which will have a resolution better than that of an optical microscope comes to mind. Such an instrument was designed by Ernst Ruska in 1930. It is called an electron microscope. It uses electric and magnetic lenses which provide focussing effect to a beam of electrons. Ruska shared the 1986 Nobel prize in physics with Gerd Binning and Heinrich Rohrer who were recognized for their design of the scanning tunneling microscope. The old dream of designing an instrument with a resolution of an angstrom good enough to see individual atoms has been realized.

The dual nature of matter, namely the fact that a beam of particles say electrons possesses both particle like and wave like attributes, and the dual nature of electro-

magnetic waves, namely the fact that a monochromatic wave behaves like a stream of particles of definite energy and momentum, appear at first sight to be mind boggling. These conflicting ideas cannot be reconciled within the framework of Newtonian mechanics and Maxwell's electrodynamics.

Banesh Hoffmann in his book 'The Strange Story of the Quantum' has written, "It is well that the reader should appreciate through personal experience the agony of the physicists of that period. They could but make the best of it, and went around with woebegone faces sadly complaining that on Mondays, Wednesdays, and Fridays they must look on light/electron as a wave; on Tuesdays, Thursdays, and Saturdays, as a particle. On Sundays they simply prayed."

As mentioned earlier only in the new mechanics of Heisenberg and Schrödinger, which surrenders a classical description of microscopic phenomenon occurring in space and time, the dual nature of matter and light appear in a natural way as two extremes of the state of a system. Heisenberg worked out that there exist a pair of variables called "conjugate variables" which cannot be measured simultaneously with arbitrary accuracy. Two such variables are position and momentum. The act of precise measurement of position of an electron is at the cost of an unavoidable vagueness of its momentum and vice-versa. When the position of an electron is known accurately, that is, when it is observed to possess particle-like attributes, a precise de Broglie wavelength cannot be associated with it, because the act of measurement of position brings in unavoidable vagueness in its momentum and hence there is a loss of its wave-like attributes. When an electron/ photon is measured to possess a definite momentum, it behaves like a wave of definite wavelength. The act of measurement of the momentum results in a consequent vagueness in the position and there is an unavoidable loss of the particle-like attributes. Therefore, electrons/ photons can be thought of either as particles, or as waves, but not both at the same time! This is the essence of Heisenberg's uncertainty principle. In this mechanics the act of observation of a system say an atomic process, can result in a discontinuous change in the physical system itself.

For going deeper into such questions we have to study quantum mechanics systematically. Quantum mechanics requires more mathematics than you have learnt for its detailed description. It is beyond the scope of our course. You will have to wait till your higher studies in physics and mathematics for a formal introduction to quantum mechanics. However, using some rules of thumb when combined with classical physics both qualitative and quantitative descriptions of simple atomic and molecular phenomena can be worked out. That is the subject of the following Chapter.

Table 12.3:

Year	Name	Citation of the Nobel prize
1901	Wilheim Conrad Rontgen	In recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him.
1902	Hendrik Antoop Lorentz and Pieter Zeeman	In recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena.
1903	Antoine Henri Becquerel	In recognition of the extraordinary services he has rendered by his discovery of spontaneous radioactivity
	Pierre Curie and Marie Sklodowska - Curie	In reconginition of the extraordinary services they have rendered by their joint researches on the radiation phenomena discovered by Professor Henri Becquerel.
1905	Philip Edward Anton Von Lenard	For his work on cathods rays.
1906	Joseph John Thomson	In recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases.
1914	Max Von Laue	For his discovery of the diffraction of Rontgen rays by crystals.
1915	William Henry Bragg and William Lawrence Bragg	For their services in the analysis of crystal structure by means of Rontgen rays.
1918	Max Planck	In recognition of the services he rendered to the advancement of physics by his discovery of energy quanta.
1921	Albert Einstein	For his services to theoretical physics and specially for his discovery of the law of photoelectric effect.
1922	Niels Bohr	For his services in the investigation of the structure of atoms and of radiation emanating from them.

1923	Robert Andrews Millikan	For his work on the elementary charge of electricity and on the photoelectric effect.
1925	James Franck and Gustav Hertz	For their discovery of the laws governing the impact of an electron upon an atom.
1927	Arthur Holly Compton	For his discovery of the effect named after him.
	Charles Thomson Rees Wilson	For his method of making the paths of electrically charged particles visible by condensation of vapour.
1928	Prince Louis Victor de Broglie	For his discovery of the wave nature of electrons.
1930	Sir Chandrasekhara Venkata Raman	For his work on the scattering of light and for the discovery of the effect named after him.
1932	Werner Heisenberg	For the creation of quantum mechanics, the application of which has, among other things led to the discovery of the allotropic forms of hydrogen.
1933	Erwin Schrödinger and Paul Adrian Maurice Dirac	For the discovery of the new productive forms of atomic theory.
1935	James Chadwick	For his discovery of the neutron.
1936	Victor Franz Hess	For his discovery of cosmic radiation.
	Carl David Anderson	For his discovery of the positron.
1937	Clinton Joseph Davisson and George Paget Thomson	For their experimental discovery of the diffraction of electrons by crystals.
1953	Enrico Fermi	For his demonstration of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons.

Summary

1. A typical arrangement for studying discharge through gases consists of a glass tube with electrodes at its two ends connected to an induction coil. The tube is connected to a vacuum pump. At atmospheric pressure, the gas in the tube is non-conducting. At low pressures, a glow (discharge) appears which shows many effects such as striations as the pressure is reduced. Below a certain low pressure, the interior of the tube becomes dark.

2. *Cathode rays* are emissions from the cathode of a discharge tube which cause fluorescence on the walls or a screen. They travel in straight lines causing shadows of obstacles in their paths. They are emitted perpendicular to the surface of the cathode and are deflected by electric and magnetic fields. Their properties are independent of the material of the cathode. Thomson identified them as a beam of negatively charged universal constituents of matter, now called electrons.

3. *Thomson's set-up for the determination of e/m of electrons*: In this method a collimated cathode ray beam emerging out of the anode passes through a region of mutually perpendicular electric and magnetic fields both normal to the beam direction. The strengths of $\mathbf{E}$ and $\mathbf{B}$ are adjusted such that the beam remains undeflected. This means $eE = evB$, where v is the speed of the electrons. If V is the accelerating voltage between the cathode and the anode, $eV = (1/2)mv^2$. These relations give,

   $$\frac{e}{m} = \frac{E^2}{2VB^2}$$

   The measured value of e/m for electrons is

   1.759×10^{11} C kg^{-1}

4. *Millikan's method for measurement of e*: A small oil droplet of radius a and mass m falling under gravity acquires a terminal speed v_g given by,

   $$6\pi\eta a v_g = mg$$

   where η is viscosity of air.

   If the droplet has charge q and a uniform electric field E is applied so that the droplet is attracted upwards and acquires a terminal speed v_e, we have

   $$6\pi\eta a v_e = qE - mg$$

   If t_g and t_e are the times taken by the droplet to cover a fixed distance D in the two cases above,

   $$q = \frac{mg\,t_g}{E}\left(\frac{1}{t_e} + \frac{1}{t_g}\right)$$

   If the same droplet acquires charge q', then

$$\frac{q'}{q} = \frac{1/t'_e + 1/t_g}{1/t_e + 1/t_g}$$

By measuring these times, Millikan found that

$$\frac{q'}{q} = \frac{n'}{n}$$

where n, n' are integers.

This establishes quantization of charge i.e., charge can occur only as integral multiple of a fundamental charge e. The measured value of e is:

$$e = 1.602 \times 19^{-19}\ \mathrm{C}$$

From the known value of e/m, we then get

$$m = 9.1 \times 10^{-31}\ \mathrm{kg}.$$

5. **Photoelectric effect:** is the emission of electrons by metals when illuminated by (ultraviolet) light.

Observed features:

(a) Photoelectric current (or number of photo electrons) is proportional to the intensity of incident light; but the *maximum kinetic energy of electrons* $(KE)_{max}$ [which equals eV where V is the cut-off voltage] is *independent* of the intensity of light.

(b) The emission of photoelectrons is almost instantaneous (time delay less than 10^{-9} s).

(c) Photoelectrons are not emitted if the frequency ν of incident light is less than the threshold frequency ν_0 characteristic of the metal, no matter how high the intensity.

(d) The maximum kinetic energy of photoelectrons and the frequency of incident light are related linearly as

$$KE_{max} = eV \propto (\nu - \nu_0)$$

Theoretical Inference: These features cannot be reconciled with classical wave picture of light in which electrons in the metal can continuously absorb energy from light. They are explained by Einstein's theory in which radiation of frequency ν consists of discrete quanta (photons) each of energy $h\nu$, and photoelectric effect is seen in terms of absorption of a photon by an electron. The minimum energy needed for a free electron in a metal to come out of its surface is called the *work function* W of the metal. Applying energy conservation to photon absorption by an electron in the metal,

$$h\nu = KE_{max} + W$$

Writing $W = h\nu_0$, we get

$$KE_{max} = h(\nu - \nu_0)$$

which is the same as feature (iv) above with the proportionality constant equal to Planck's constant h. Millikan's first accurate measurements confirmed this equation and obtained $h = 6.6 \times 10^{-34}$ Js in agreement with the value of h obtained from blackbody radiation measurements.

6. *A photocell* employs photoelectric effect to convert a change in intensity of illumination into a change in electric current. Photocells are used to operate controls, and as light meters(in camera, for example)

7. *de Broglie's hypothesis*: A beam of material particles each travelling with momentum p will manifest wave character with wavelength $\lambda = h/p$, where h is planck's constant.

   A macroscopic object has exceedingly small λ and so does not show wave like property.

   Davisson and Germer experiment used a beam of electrons, each with energy 54 eV. The corresponding de Broglie wavelength λ is equal to 1.66 Å. They studied diffraction of this beam by a nickel crystal, in which the smallest separation d between atoms is $d = 0.914$ Å. The diffraction peak was found at an angle of 65° from the axis of the incident beam. Using Bragg's relation $2d\sin\theta = \lambda$, λ is found to be 1.65 Å in close agreement with that predicted by the de Broglie relation. The experiment confirmed the wave nature of matter.

8. Electron microscope uses this wave nature of electrons to provide high resolving power.

9. Wave-particle duality of light and matter cannot be understood in classical physics but can be dealt with in modern quantum mechanics.

Exercises

12.1 Outline the principle of Thomson's method for the measurement of e/m.

12.2 In a Thomson set-up, the electric field strength $E = 30\text{Vcm}^{-1}$ and the magnetic field strength $B = 6.0\text{G}$. What is the speed of the electron that goes undeflected in the common region of the two fields?

12.3 What is the acceleration of an electron in an electric field of magnitude 50Vcm^{-1} given that its $e/m = 1.76 \times 10^{11}\text{Ckg}^{-1}$?

12.4 What is the magnitude of acceleration of an electron of speed $2.5 \times 10^6\text{ms}^{-1}$ in a magnetic field of 2.0 G given that its $e/m = 1.76 \times 10^{11}\text{Ckg}^{-1}$?

12.5 In a Thomson set-up for the determination of e/m, electrons accelerated by 2.5 kV enter the region of crossed electric and magnetic fields of strengths $3.6 \times 10^4\text{Vm}^{-1}$ and $1.2 \times 10^{-3}\text{T}$ respectively and go through undeflected. Determine the e/m of an electron.

12.6 In an experiment to determine e/m using Thomson's method, electrons from the cathode accelerate through a potential difference of 1.5 kV. The beam coming out of a hole in the anode is collimated further and then made to enter a region of crossed electric and magnetic fields both perpendicular to the beam direction. Two parallel plates 2.0 cm apart held at a potential difference of 400 V provide the electric field, while the magnetic field is produced by a Helmholtz coils arrangement. The current in the coils is so adjusted that the beam hits the fluorescent screen at the end of the tube *undeflected*. The magnetic field due to the coils is calculated to be 8.6×10^{-4} T. Determine the charge to mass ratio of an electron.

12.7 Outline the principle of Millikan's method for the determination of fundamental charge.

12.8 In a Millikan's experiment an oil drop of radius 1.5×10^{-6} m and density 890 kg m^{-3} is held stationary between two condenser plates 1.2 cm apart and kept at a potential difference of 2.3 kV. How many excess electrons are carried by the drop? ($e = 1.6 \times 10^{-19}\text{C}$; $g = 9.8\text{ms}^{-2}$) Ignore buoyancy due to air.

12.9 A charged oil drop falls under gravity with a terminal speed v. The drop is held stationary by applying suitable electric field in a Millikan's set-up and is found to carry 2 excess electrons. Suddenly the drop is observed to move upwards with speed v. Guess what has happened.

12.10 State some important observed features of photoelectric effect which cannot be reconciled with the classical wave picture of light and its absoption by electrons.

12.11 Explain how Einstein's quantum view of the photoelectric effect agrees with the important observed features of this process.

12.12 The photoelectric cut-off voltage in a certain experiment is 1.5 V. What is the maximum kinetic energy of photoelectrons emitted? $e = 1.6 \times 10^{-19}$C.

12.13 In an experiment on photoelectric effect, the slope of the cut-off voltage versus frequency of incident light is found to be 4.12×10^{-15} Vs. Given $e = 1.60 \times 10^{-19}$C, estimate the value of Planck's constant.

12.14 The threshold frequency for a certain metal is $3.3. \times 10^{14}$ Hz. If light of frequency 8.2×10^{14}Hz is incident on the metal, predict the cut-off voltage for photoelectric emission. (Use the known values of h and e).

12.15 The work function for a certain metal is 4.2 eV. Will this metal give photoelectric emission for incident radiation of wavelength 330 nm? (Use the known values of e,h and c)

12.16 For each statement below, state with reasons if it is true or false.

(a) The miximum kinetic energy of photoelectrons depends on the frequency of radiation and material of the photocell.

(b) If the intensity of incident radiation is doubled, the maximum kinetic energy of phtoelectrons is also doubled.

(c) The cut-off voltage in a photoelectric experiment has a linear dependence on the frequency of incident radiation.

(d) The photon picture of electromagnetic radiation predicts nearly instantaneous photoelectric emission.

12.17 A photon of wavelength 1.50×10^{-10}m is scattered by a free electron initially at rest. If the final kinetic energy of the electron is 2.60×10^{-17}J, what is the wavelength of the scattered photon? (Use the known values of c and h)

12.18 What is the change in wavelength of an X-ray photon in a Compton scattering experiment with electrons if the scattered photon makes an angle of 60^0 with the incident photon direction? Use the knowr values of h,c and mass of the electron.

12.19 State de Broglie hypothesis on the wave nature of matter. Obtain the de Broglie wavelength of an electron of kinetic energy 100eV. Mass of electron $= 9.1 \times 10^{-31}$ kg, $e = 1.6 \times 10^{-19}$C, $h = 6.63 \times 10^{-34}$Js.

12.20 Outline schematically the experiment of Davisson and Germer to demonstrate the wave nature of matter.

12.21 For what kinetic energy of a neutron will the associated de Broglie wavelength be 1.40×10^{-10}m? Mass

of a neutron = 1.675×10^{-27} kg, $h = 6.63 \times 10^{-34}$Js.

12.22 Show that the wavelength of electromagnetic radiation is equal to de Broglie wavelength of its quantum (photon).

12.23 For a given kinetic energy, which of the following has the smallest de Broglie wavelength; electron, proton, α - particle?

Additional Exercises

12.24 (a) Estimate the speed with which electrons emitted from a heated cathode of an evacuated tube impinge on the anode maintained at a potential difference of 500 V with respect to the cathode. Ignore the small initial speeds of the electrons. The 'specific charge' of the electron i.e. its e/m is given to be 1.76×10^{11}Ckg$-$1.

(b) Use the same formula you employ in (a) to obtain electron speed for an anode potential of 10 MV. Do you see what is wrong? In what way is the formula to be modified?

12.25 (a) A monoenergetic electron beam with electron speed of $5.20 \times 10^{6}\text{ms}^{-1}$ is subject to a magnetic field of 1.30×10^{-4}T normal to the beam velocity. What is the radius of the circle traced by the beam, given e/m for electron equals $1.76 \times 10^{11}\text{Ckg}^{-1}$.

(b) Is the formula you employ in (a) valid for calculating radius of the path of a 20 MeV electron beam? If not, in what way is it modified?

Note: Exercises 12.24(b) and 12.25(b) take you to relativistic mechanics which is beyond the scope of this book. They have been inserted here simply to emphasize the point that the formulas you use in part (a) of the exercises are not valid at very high speeds or energies. See answers at the end to know what 'very high speed or energy' means.

12.26 In a Thomson's set-up for determining e/m, the same high tension dc supply provides potential to the anode of the accelerating column, as also to the positive deflecting plate in the region of crossed fields. If the supply voltage is doubled, by what factor should the magnetic field be increased to keep the electron beam undeflected?

12.27 The deflecting plates in a Thomson's set-up are 5.0 cm long, and 1.5 cm apart. The plates are maintained at a potential difference of 240 V. Electrons accelerated to an energy of 2.0 keV enter from one edge of the plates midway in a direction parallel to the plates.

(a) What is the deflection at the other edge of the plates?

(b) At what distance from the undeflected position of the screen does the beam strike if the screen is 30 cm away from the other edge of the plates?

12.28 In a Thomson's set-up for determination of e/m, a uniform electric

field $E = 24.0\text{kVm}^{-1}$ set up between two parallel plates of length 6.0 cm produces a deflection of 10.9 cm on the fluorescent screen. A magnetic field is then switched on and adjusted to the value $B = 8.0 \times 10^{-4}\text{T}$ to restore the beam to its undeflected position. The distance of the screen from the centre of the plates is 40.0 cm. Determine e/m from the data.

12.29 An electron gun with its anode at a potential of 100 V fires out electrons in a spherical bulb containing hydrogen gas at low pressure ($\sim 10^{-2}$mm of Hg). A magnetic field of $2.83 \times 10^{-4}\text{T}$ curves the path of the electrons in a circular orbit of radius 12.0 cm. (The path can be viewed because the gas ions in the path focus the beam by attracting electrons, and emitting light by electron capture; this method is known as the 'fine beam tube' method.) Determine e/m from the data.

12.30 In a Millikan's oil drop experiment, a charged oil drop of mass density 880 kg m^{-3} is held stationary between two parallel plates 6.00 mm apart held at a potential difference of 103V. When the electric field is switched off, the drop is observed to fall a distance of 2.00 mm in 35.7 s.

(a) What is the radius of the drop?

(b) Estimate the charge of the drop. How may excess electrons does it carry? (The upper plate in the experiment is at a higher potential). (Viscosity of air $= 1.80 \times 10^{-5}\text{Nsm}^{-2}$, $g = 9.81\text{ms}^{-2}$; density of air $=$ 1.29 kgm^{-3}).

12.31 (Millikan's experiment not only measured charge of an electron; it also established a fundamental fact of nature: charge is quantized i.e. charge comes in multiples of a basic unit namely the electronic charge e. Make sure you understand this point by going through the following excercise:) In a Millikan's oil drop set-up, an oil drop falls with a terminal speed v_0 in the absence of any electric field. When a fixed electric field is switched on, and a radioactive source is kept in the environment, it is found that the same oil drop, when viewed for long, shows up different terminal speeds v_1, v_2, $v_3 \ldots$

(a) What causes the drop to change its terminal speed in the same electric field? (Assume size and mass of the drop remain unchanged.)

(b) What key observation on the different terminal speeds of the drop suggests charge quantization?

12.32 In a variant of the Millikan's oil drop set-up, an oil drop whose radius is measured by a separate observation to be 1.0×10^{-6}m, falls down in the absence of any electric field with a certain terminal velocity. When a horizontal electric field is set up by means of two parallel vertical plates held 10 mm apart at a potential difference of 1500 V, the drop is seen to fall steadily at an angle of 63°

with the vertical. The density of the oil used is 900kgm^{-3}. Estimate the charge on the drop.

12.33 (a) An x-ray tube produces a continuous spectrum of radiation with its short wavelength end at 0.45 Å. What is the maximum energy of a photon in the radiation?

(b) From your answer to (a), guess what order of accelerating voltage (for electrons) is required in such a tube?

12.34 In an accelarator experiment on high energy collisions of electrons with positrons, a certain event is interpreted as annihilation of an electron-positron pair of total energy 10.2 BeV into two γ-rays of equal energy. What is the wavelength associated with each γ-ray? ($1\text{BeV} = 10^9\text{eV}$)

12.35 Estimating the following two numbers should be interesting. The first number will tell you why radio engineers do not need to worry much about photons! The second number tells you why our eye can never 'count photons', even in barely detectable light. (i) The number of photons emitted per second by a Medium wave transmitter of 10kW power emitting radiowaves of wavelength 500 m. (ii) The number of photons entering the pupil of our eye per second corresponding to the minimum intensity of white light that we humans can percieve ($\sim 10^{-10}\text{Wm}^{-2}$). Take the area of the pupil to be about 0.4 cm^2, and the average frequency of white light to be about 6×10^{14} Hz.

12.36 Ultraviolet light of wavelength 2271 Å from a 100 W mercury source irradiates a photocell made of molybdenum metal. If the stopping potential is −1.3V, estimate the work function of the metal. How would the photocell respond to a high intensity ($\sim 10^5\text{Wm}^{-2}$) red light of wavelength 6328 Å produced by a He - Ne laser?

12.37 Monochromatic radiation of wavelength 640.2 nm ($1\text{nm} = 10^{-9}$ m) from a neon lamp irradiates photosensitive material made of caesium on tungsten. The stopping voltage is measured to be 0.54 V. The source is replaced by an iron source and its 427.2 nm line irradiates the same photocell. Predict the new stopping voltage.

12.38 A mercury lamp is a convenient source for studying frequency dependence of photoelectric emission, since it gives a number of spectral lines ranging from the UV to the red end of the visible spectrum. In our experiment with rubidium photocell, the following lines from a mercury source were used:

$\lambda_1 = 3650$ Å, $\lambda_2 = 4047$ Å, $\lambda_3 = 4358$ Å, $\lambda_4 = 5461$ Å, $\lambda_5 = 6907$ Å

The stopping voltages, respectively, were measured to be:

$V_0^1 = 1.28\text{V}$, $V_0^2 = 0.95\text{V}$, $V_0^3 = 0.74\text{V}$, $V_0^4 = 0.16\text{V}$, $V_0^5 = 0\text{V}$

(i) Determine the value of Planck's constant h.

(ii) Estimate the threshold frequency and work function for the material.

Note: You will notice that to get h from the data, you will need to know e (which you can take to be 1.6×10^{-19}C). Experiments of this kind on Na. Li, K etc. were performed by Millikan who, using his own value of e (from the oil drop experiment) confirmed Einstein's photoelectric equation and at the same time gave an independent estimate of the value of h.

12.39 The work function for the following metals is given:

Na: 1.92eV; K: 2.15eV; Mo: 4.17eV; Ni: 5.0eV. Which of these metals will not give p.e. emission for a radiation of wavelength 3300Åfrom a He-Cd laser placed 1 m away from the photocell? What happens if the laser is brought nearer and placed 50 cm away?

12.40 Light of intensity 10^{-5}Wm^{-2} falls on a sodium photocell of surface area 2 cm^2. Assuming that the top 5 layers of sodium absorb the incident energy, estimate time required for photoelectric emission in the wave picture of radiation. The work function for the metal is given to be about 2eV. What is the implication of your answer?

12.41 (a) Show that a free electron at rest cannot absorb a photon and thereby acquire kinetic energy equal to the energy of the photon. Would the condusion change if the free electron was moving with a constant velocity? (b) if the absorption of a photon by a free electron is ruled out as proved in (a) above, how does photoelectric emission take place at all?

Note: It will be worth going through the answer to this excercise given at the end. Though some relations used there are relativistic (and beyond the scope of this book), the basic idea that both energy and momentum should be conserved in a process is well-known to you from class XI. Also, it is important to be clear about the point in (b).

12.42 In an experiment on photoelectric emission by γ-rays on platinum, the energy distribution of photoelectrons exibits peaks at a number of discrete energies: 270 keV, 339 keV and 354 keV. The binding energies of K,L and M shells in platinum are known to be 77 keV, 13 keV and 3.5keV aproximately. What is the wavelength of the γ-rays with which the data are consistent?

12.43 An X-ray pulse is sent through a section of Wilson cloud chamber containing a supersaturated gas, and tracks of photoelectrons ejected from the gaseous atoms are observed. Two groups of tracks of lengths 1.40 cm and 2.02 cm are noted. If the range-energy relation for the cloud chamber is given by $R = \alpha E$ with $\alpha = 1$ cm/keV, obtain the binding energies of the two levels from which electrons are emitted. (Wavelength of the X-rays pulse = 4.9 Å)

Note: Exercises 12.42 and 12.43

take you away from the typical scenario of photoelectric effect that you have learnt in this chapter. Actually, like visible and UV radiation, x - rays and γ-rays also cause p.e. emission. But their photons have much greater energies, and so can eject electrons of much greater binding energies than those ejected by visible or UV. They can eject electrons from the inner shells of individual atoms where the energies are discrete. The energy distribution of electrons emitted by x-ray or γ-ray photoelectric effect may, consequently, exhibit sharp peaks at discrete energies (at the lower end of the energy spectrum).

12.44 Crystal diffraction experiments can be performed using x-rays, or electrons accelerated through appropriate voltage. Which probe has greater energy? An x-ray photon or the electron? (For quantitative comparison, take the wavelength of the probe equal to 1 Å, which is of the order of inter-atomic spacing in the lattice) ($m_e = 9.11 \times 10^{-37}$kg.)

12.45 (a) Obtain the de Broglie wavelength of a neutron of kinetic energy 150 eV. As you have seen in 12.44, an electron beam of this energy is suitable for crystal diffraction experiments. Would a neutron beam of the same energy be equally suitable? Explain. ($m_n = 1.675 \times 10^{-27}$kg)

(b) Obtain the de Broglie wavelength associated with thermal neutrons at room temperature (27°C). Hence explain why a fast neutron beam needs to be thermalized with the environment before it can be used for neutron diffraction experiments.

12.46 An electron microscope uses electrons accelerated by a voltage of 50 kV. Determine the de Broglie wavelength associated with the electrons. If other factors (such as numerical aperture, etc) are taken to be roughly the same, how does the resolving power of an electron microscope compare with that of an optical microscope which uses yellow light?

12.47 The wavelength of a probe is roughly a measure of the size of a structure that it can probe in some detail. The quark structure of protons and neutrons appears at the minute length scale of 10^{-15}m or less. This structure was first probed in early 1970's using high energy electron beams produced by a linear Accelerator at Stanford, USA. Guess what might have been the order of energy of these electron beams. (Rest mass energy of electron = 0.511 MeV).

12.48 The extent of localization of a particle is determined roughly by its de Broglie wavelength. If an electron is localized within the nucleus (of size about 10^{-14}m) of an atom, what is its energy? Compare this energy with the typical binding energies (of the order of a few MeV) in a nucleus, and hence argue why electrons can not reside in a nucleus.

12.49 Find the typical de Broglie wavelength associated with a He atom in helium gas at room temperature (27°C) and 1 atm. pressure; and compare it with the mean separation between two atoms under these conditions.

12.50 Compute the typical de Broglie wavelength of an electron in a metal at 27°C and compare it with the mean separation between two electrons in a metal which is given to be about 2×10^{-10}m.

Note: Exercises 12.49 and 12.50 reveal that while the wave-packets associated with gaseous molecules under ordinary conditions are non-overlapping, the electron wave packets in a metal strongly overlap with one another. This suggests that whereas molecules in an ordinary gas can be distinguished apart, electrons in a metal can not be distinguished apart from one another. This *indistinguishibility* has many fundamental implications which you will explore in more advanced physics courses.

12.51 Answer the following questions:

(a) Quarks inside protons and neutrons are thought to carry fractional charges $(+2/3e, -1/3)$. Why do they not show up in Millikan's oil drop experiment?

(b) Why need the oil drops of Millikan's experiment be of such microscopic sizes? Why can not we experiment with much bigger drops?

(c) Stokes' formula for viscous drag is not really valid for oil drops of extremely minute sizes. Why not?

(d) What is so special about the combination e/m? Why do we not simply talk of e and m separately?

(e) Why should gases be insulators at ordinary pressures and start conducting at very low pressures?

(f) Every metal has a definite work function. Why do photoelectrons not come out all wilth the same energy if incident radiation is monochromatic? Why is there an energy distribution of photoelectrons?

(g) The energy and momentum of an electron are related to the frequency and wavelength of the associated matter wave by the relations:

$$E = hv, \quad P = \frac{h}{\lambda}$$

But while the value of λ is physically significant, the value of ν (and therefore the value of the phase speed $\nu\lambda$) has no physical significance. Why?

Thomson, Sir John Joseph (1856-1940) British scientist who investigated the nature of cathode rays. He discovered the electron by his ingenious experiments on discharge of electricity through gases.

de Broglie, Prince Louise-Victor (1892-1987) French physicist who put forth revolutionary idea of wavenature of matter. This idea was developed by Erwin Schrödinger into a full-fledged theory of quantum mechanics commonly known as wave mechanics.

CHAPTER **13**

Atoms, Nuclei and Molecules

13.1 Introduction

In this Chapter we shall study the structure of the atom. In the nineteenth century, enough evidence had accumulated in favour of the hypothesis that each element has its distincive atom. Experiments described in the last Chapter have revealed that atoms of different elements contain electrons which are completely identical.

The electrical neutrality of atoms is evident even at the macroscopic level. A piece of glass is electrically neutral. It acquires a net electric charge when rubbed against a silk handkerchief. We also know from the last Chapter that each electron carries negative electric charge equal in magnitude to

the fundamental charge 1.6×10^{-19}C and is thousands of times lighter than any atom. This clearly indicates that atoms must contain positive charge. Because if atoms are strictly electrically neutral, the amount of positive charge in each atom has to be equal to the number of electrons it contains multiplied by the fundamental charge.

As of now 105 elements are known to man. It is therefore reasonable to suppose that atoms of different elements differ from each other because they contain different numbers of electrons and positive charges. How are the electrons and positive charges distributed inside the atom? Does the positive charge reside on particles or it is uniformly distributed inside the atom? In other words, what is the structure of the atom?

In 1898, J.J. Thomson proposed that the atom is basically a spherical cloud of positive charges with electrons embedded in it like seeds in a watermelon. At that time, electrons had been identified, but not much was known about the positive charge. The location of electrons in the charge cloud was to be determined theoretically by imposing the requirement that the system should be stable. Slight disturbance of the atom were expected to cause oscillations of the electrons about their equilibrium positions such that light of definite frequencies could be radiated by the system. This model was taken up seriously although electrostatically it is an unstable system.

Ernest Rutherford (1871-1937), a former student of Thomson, was engaged at Manchester in epoch-making experiments using high energy α (alpha)-particles emitted by some radioactive elements. Rutherford identified alpha-particles with helium atoms which had lost both of their electrons. They, therefore, carry two units of positive charge and are about 8,000 times heavier than electrons. The speed and therefore the energy of the alpha-particles are characteristics of their radioactive parent. By shooting alpha-particles at thin metal foils and observing the change in their paths, if any, Rutherford hoped to investigate the force experienced by them during their passage through the atoms of the metal. He observed that although most of the alpha-particles travelled through the thin metal foil undeviated, some suffered a large change in direction. This indicates that on some occasions an alpha-particle experiences a large force.

The deflection of the positively charged alpha-particle in passing through the atom can be due to two causes - (i) attraction by the negatively charged electrons and (ii) repulsion by the positive charge in the atom.

The hypothesis that the large change in direction of alpha-particles is due to attraction by electrons in atoms is more impalusible than the explanation that a fast moving bullet rebounded because it hit a mosquito. Further, theoretical estimates showed that the positive charge in Thomson's atom is spread too thin to exert the strong forces experienced by some of the alpha-particles. At the suggestion of Rutherford, in 1911, H. Geiger and E. Marsden performed some classic experiments. In one of their experiments, 5.5 MeV alpha-particles (velocity = 1.63×10^7 m/s) from a ${}^{214}_{83}$Bi source were scattered from a thin gold foil of thickness 2.1×10^{-7}m. The experiments of Geiger and Marsden supplied information about the nucleus.

These observations led to the birth of Rutherford's planetary model of the atom which is also called the nuclear model of the atom. In this model the mass of the atom and all its positive charge is concentrated in a tiny nucleus and electrons revolve around the nucleus like planets around the sun. It may be noted that in 1901 Jean Perrin had published a paper on "The nucleo-planetary

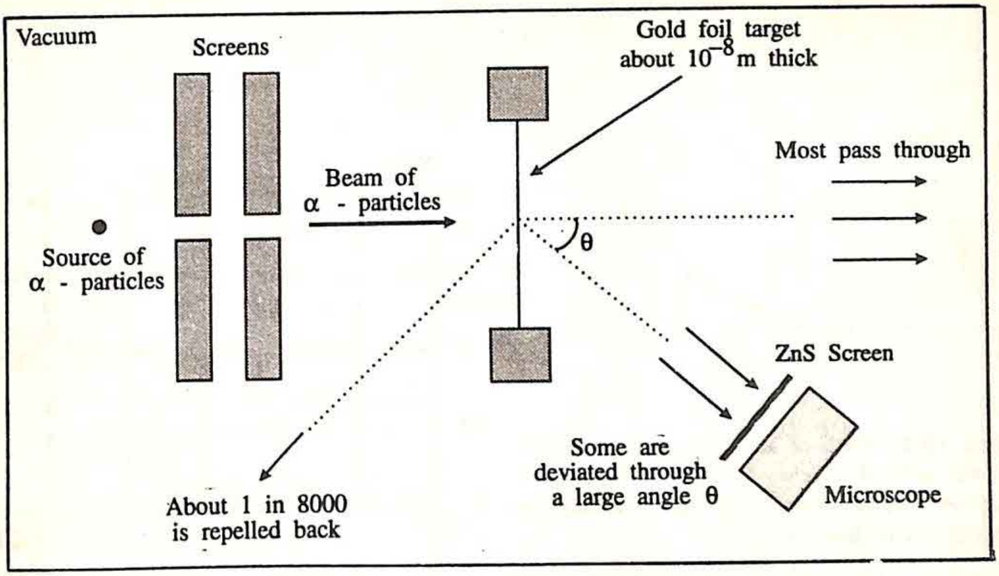


Figure 13.1: Schematic arrangement of the Geiger-Marsden experiment.

structure of the atom" and independently in 1903 a similar model was proposed by Nagaoka. But, it was Rutherford who established such a model on a solid experimental foundation.

We discuss next the arrangement of the experiment and analysis of the observations on alpha-particle scattering based on Rutherford's nuclear model of the atom.

13.2 Alpha-particle scattering and Rutherford's model of atom

The arrangement of the Geiger-Marsden experiment is shown in Fig. 13.1. Alpha-particles emitted by a $^{214}_{83}$Bi radioactive source are collimated into a narrow beam by their passage through lead bricks. The beam was allowed to fall on a thin gold foil of thickness 2.1×10^{-7}m. The scattered alpha-particles were observed through a rotatable detector consisting of a zinc sulfide screen and a microscope. The alpha-particles on striking the screen produced bright flashes, or scintillations, which could be observed and counted at different angles from the direction of the incident beam.

The graph of the total number of alpha-particles scattered at different angles agrees with the calculations made by Rutherford using the nuclear model of the atom. It is assumed that the gold foil is so thin that alpha particles will suffer not more than one scattering during their passage through it. Therefore, a computation of the trajectory of an alpha-particle in its interaction with a single nucleus is enough. We assume that alpha particles are nuclei of helium atoms. Therefore, they carry two units, $2e$, of positive charge and have the mass of the helium atom. The charge of the gold nucleus is Ze, with $Z = 79$. Since the nucleus of a gold atom is about 50 times heavier than an alpha-particle, it is reasonable to assume that it remains at rest throughout the scattering process. Under these assumptions the path or trajectory of an alpha-particle can be computed using Newton's second law of motion and Coulomb's law for the electrostatic force of repulsion between an alpha particle and the positively charged nucleus.

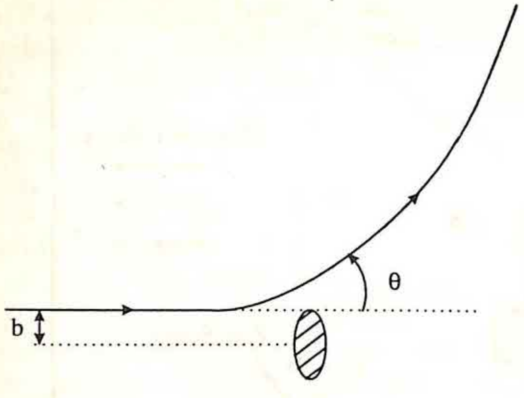


Figure 13.2: Path of an alpha-particle in the Coulomb field of a heavy nucleus. The impact parameter b and the scattering angle θ are defined as shown in the diagram.

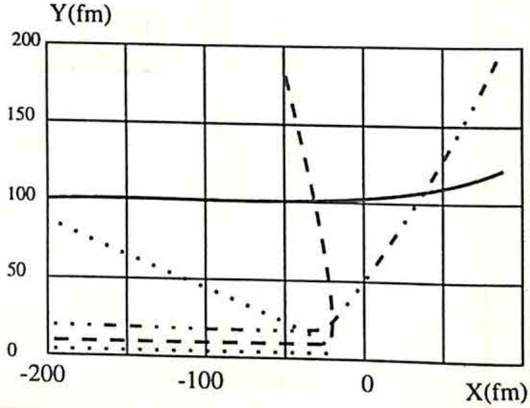


Figure 13.3: Numerically calculated trajectories of α-particles, moving with speed of 1.63 $\times$ 10^7 m/s far away from the nucleus, in the Coulomb field of a gold nucleus. The gold nucleus is located at the origin of the coordinate system. The values of the impact parameters are 2.5 fm, 10 fm, 20 fm and 100 fm.

The magnitude of this force is given by the expression

$$F = \frac{1}{4\pi\epsilon_0} \frac{(2e)(Ze)}{r^2}\ , \tag{13.1}$$

where Z is the atomic number of the nucleus and r is the distance between the α-particle and the nucleus. The force is directed along the line joining the two nuclei. The magnitude and direction of the force on an alpha particle continuously change as it approaches the nucleus and when it recedes away from it.

The scattering of an alpha particle from a nucleus depends on its distance of closest approach to the nucleus or an equivalent length called the *impact parameter*, b, as defined in Fig. 13.2. You may note that it is the perpendicular distance of the velocity vector of the alpha particle from the centre of the nucleus when it is far away from the atom. In Fig. 13.3 we have shown the graphs of numerically calculated trajectories for the Geiger-Marsden experiment for four values of the impact parameters. The values of these impact parameters are 2.5 fm, 10 fm, 20 fm and 100 fm (1 fm = 10^{-15}m). You can easily note that for large impact parameters, the force experienced by the alpha particle is weak because of its inverse square law character. For a large impact parameter an alpha-particle will go undeviated and for a small impact parameter, it will suffer large scattering. In fact, for the case of head-on collision the impact parameter is zero; an alpha-particle will rebound like a ball thrown against a wall.

Example 13.1: What is the distance of closest approach to the nucleus of an alpha-particle which undergoes scattering by 180° in the Geiger-Marsden experiment described above?

Answer: For this case, when the alpha-particle is closest to the nucleus, it comes to rest and its initial kinetic energy is completely converted into potential energy. We

assume that this takes place when the alpha particle is outside the nucleus. This condition is expressed by the relation

$$\frac{1}{2}mv_i^2 = \frac{1}{4\pi\epsilon_0}\frac{(2e)(Ze)}{r_0},$$

where v_i is the initial speed, Z is the atomic number of the nucleus, m is the mass of the alpha particle and r_0 is the distance of closest approach of the alpha-particle to the nucleus.

For the Geiger-Marsden experiment

$$\begin{aligned}\frac{1}{2}mv_i^2 &= 5.5\text{MeV}\\ &= 5.5\times10^6\times1.6\times10^{-19}\text{J}\\ &= 8.8\times10^{-13}\text{J}\\ Z &= 79\end{aligned}$$

Also,

$$\frac{1}{4\pi\epsilon_0} = 9.0\times10^9\ \text{Nm}^2\text{C}^{-2} \text{ and}$$

$$e = 1.6\times10^{-19}\ \text{C}$$

Therefore,

$$\begin{aligned}r_0 &= \frac{9.0\times10^9\times2\times79\times(1.6\times10^{-19})^2}{8.8\times10^{-13}}\\ &= 4.13\times10^{-14}\text{m}\\ &= 41.3\ \text{fm}\end{aligned}$$

The radius of the gold nucleus is, therefore, less than 4.1×10^{-14}m, well under (1/10,000) of the radius of the atom!

Some important features of scattering from a repulsive $1/r$ potential can be easily simulated using an analogue experiment based on the gravitational potential energy. You know that the gravitational potential energy of a particle of mass m at a height h above a laboratory table is mgh. Let us plot the graph of the function $h = 1/r$. It is as shown in Fig. 13.4. This curve when rotated about the y-axis describes a 3-dimensional object. A sheet of metal cast in the shape of the surface of revolution generated by the curve $h = 1/r$, simulates a $1/r$ potential and, therefore, can be used for studying scattering from a $1/r^2$ repulsive force (Fig. 13.5). By sliding a steel ball from a given height as shown in Fig. 13.6, an alpha-particle moving with given speed and impact parameter can be simulated. When the ball reaches the curved surface, it experiences a force field which varies as $1/r^2$. For different impact parameters the change in direction of the ball as it leaves the curved surface reveals the nature of the force field.

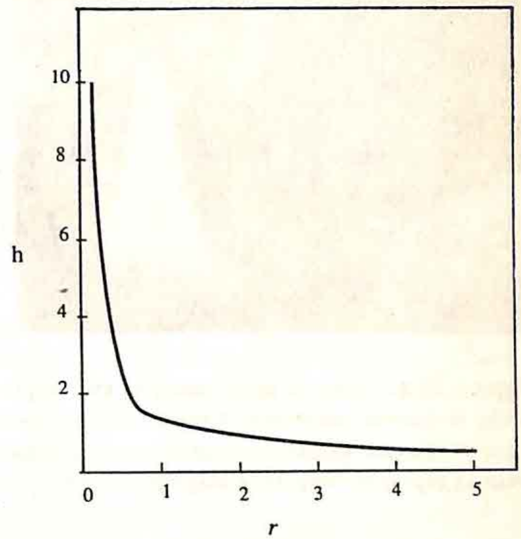


Figure 13.4: Graph of $h = \frac{1}{r}$ as a function of r.

For the analogue experiment, the graph of the impact parameter as a function of $\cot(\theta/2)$, where θ is the scattering angle is shown in Fig. 13.7. The experimental data for the variation of scattering angle as a function of impact parameter has been given in Table 13.1. Within experimental errors it is a straight line. Rutherford had analytically calculated the relation between the impact parameter b and the scattering angle θ. It is

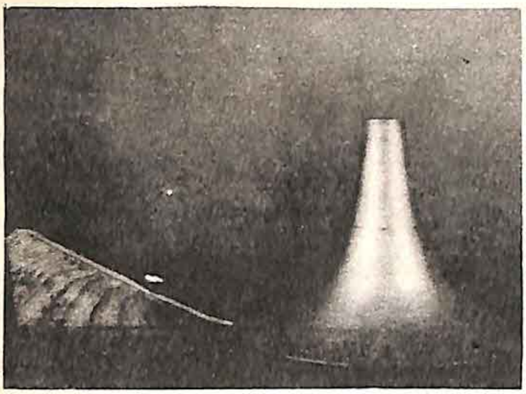


Figure 13.5: A metal sheet casted in the shape of the surface of revolution of the $1/r$ curve. Because of the gravitational potential energy it simulates a repulsive force field varying as $1/r^2$.

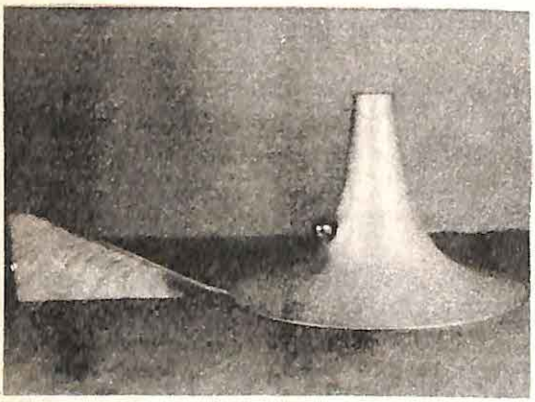


Figure 13.6: Experimental arrangement for studying the scattering from a $1/r^2$ force field using steel balls.

$$b = \frac{Ze^2 \cot(\theta/2)}{4\pi\epsilon_0 \left(\frac{1}{2}mv_i^2\right)} \tag{13.2}$$

The relation, Eq. (13.2) shows clearly that an α particle close to the nucleus (small impact parameter b) will have a large deflection ($\theta \simeq \pi$) whereas an α particle far away (large b) will have a small deflection ($\theta \simeq 0$).

Table 13.1:

b cm	θ Deg.	$\theta/2$ Deg.	cot $(\theta/2)$
1	137	68.5	0.39
2	110	55	0.70
3	88	44	1.04
4	54	27	1.96
5	44	22	2.47
6	37	18.5	2.98
7	32	16	3.48
8	27	13.5	4.16
9	25	12.5	4.5

We also see that if the kinetic energy of the α particle is large, b can be small. A given beam of α particles incident on a collection of nuclei has a distribution of impact parameters b, so that the beam is scattered in various directions with different probabilities. Knowing the probability of different values of b, and the relation Eq. (13.2) between b and θ, the relative number of particles scattered in different directions can be calculated. Geiger and Marsden found that their experiments gave values close to those expected from such a calculation. This verifies in detail the idea of a small nucleus where all positive charge *and* mass is concentrated.

The relation, Eq.(13.2) between b and θ depends on the nature of the force law. If the nucleus is a strict point charge, it holds no matter how small b is. But if the force law between the nucleus and the α particle changes as the α particle comes very close (as will happen if the nucleus has a finite size), this relation will change, and so will the angular distribution (i.e. distribution in θ) of the scattered particles. Actually, this is how the first estimates of nuclear sizes ($\sim 10^{-15}$ m) were made. In general, it is clear that measuring the angular distribution of particles scattered by a target is a powerful way of finding out the nature of forces between the scatterer and the target. Such experi-

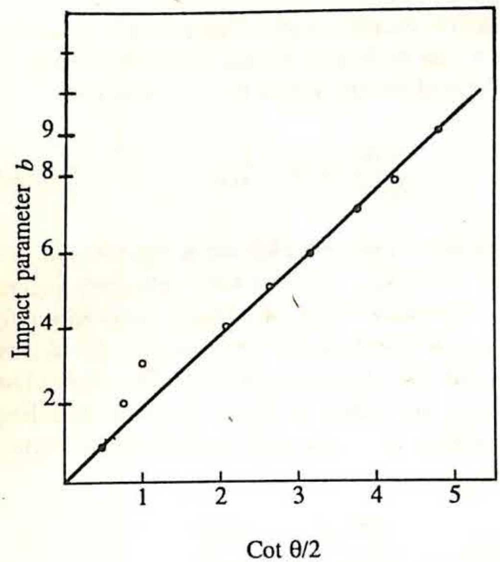


Figure 13.7: Graph of impact parameters b versus cot $(\theta/2)$.

ments are routinely performed in high energy physics, and the forces between particles have been probed to distances as small as 10^{-18}m!

In 1913, Niels Bohr used Rutherford's nuclear model of the atom for explaining the spectrum of the hydrogen atom. The simplest atom is that of hydrogen. It has a single electron, which is bound to its nucleus by the attractive Coulomb force. According to Newtonian mechanics, the electron in the hydrogen atom has to revolve in an elliptic orbit fixed by its energy and angular momentum. Because, if at any instant the electron is at rest it will be sucked in by the nucleus within no time. It turns out that orbital motion under the centripetal force is sufficient to provide stability to the planets but is not enough for the hydrogen atom! An orbiting electron is continuously under acceleration and must radiate energy as electromagnetic waves (Chapter 9, Section 9.4). As it radiates, the electron's orbit will continuously shrink spirally into the nucleus within 10^{-8}s. But we know that the hydrogen atom is stable and has a characteristic line spectrum when excited. These facts are evidently in conflict with classical physics, according to which the hydrogen atom is not stable and should give out a continuous spectrum of electromagnetic radiation!

These paradoxes were resolved by Niels Bohr. He used Rutherford's nuclear model of the atom but took a leap in physics which carried him far beyond the domain of classical physics. He laid the foundation of the quantum theory by postulating the existence of special orbits in which electrons do not radiate (for reasons not clear at the time). The new physics which supported this hypothesis was discovered a decade later by Schrödinger and Heisenberg.

The postulate on the existence of special orbits ushered in discretisation of the energy of hydrogen atom. By making another far-reaching postulate Bohr connected the transitions of hydrogen atom between a pair of allowed states with the process of emission of a photon. The frequency of the photon emitted by an atom when it makes a transition from a higher energy state to a lower energy state is determined by energy conservation. Bohr thus explained the experimentally observed spectrum of the hydrogen atom in terms of fundamental constants such as charge and mass of electron and the Planck constant. Bohr's calculation of the Rydberg constant and the spectrum of hydrogen is given in Section 13.4.

13.3 Energy quantisation

The concept of discretisation or the quantisation of physical quantities also exists in the

classical physics. In Chapter 13 of the class XI text book we saw that a stretched string or a membrane can oscillate only in definite stationary states whose fequencies are determined by elastic properties and boundary conditions. Because both the ends of a stretched string clamped at its ends must remain forever at rest, wavelengths Λ_{cc} of the allowed stationary states are related to the length L of the string. In Eq. (13.30) of the class XI textbook the allowed wavelengths have been shown to be

$$\Lambda_{cc} = \{\lambda = \frac{2L}{n}; \text{ where } n = 1, 2, 3, 4, .\} \tag{13.3}$$

The easiest way to see discretisation in quantum physics is to use the de Broglie hypothesis, namely, to recognize that particles are wavelike. We then postulate that just like a vibrating string, a free particle of mass m confined to a line of length L can have only those values of the momentum for which the de Broglie wavelength belongs to the set Λ_{cc}. Then the magnitude of its momentum is quantized.

Using

$$p = \frac{h}{\lambda} \text{ and} \lambda = \frac{2L}{n}$$

this gives

$$p = \frac{h}{(2L/n)} = \frac{nh}{2L}, \; n = 1, 2, 3, 4, \ldots \tag{13.4}$$

The energy of a free particle is given in terms of its momentum by the relation

$$E = \frac{p^2}{2m} \tag{13.5a}$$

Substituting for p the values given in Eq. (13.4) gives a simple result of quantum mechanics, namely that a free particle confined to a line of length L can have only discrete values of energy given by the relation

$$E_n = \frac{n^2h^2}{8mL^2}, n = 1, 2, 3, \ldots \tag{13.5b}$$

The parameter n which takes the integer values, 1,2,3,4, ..., labels the stationary states in ascending order of energy. According to Bohr, a quantum system can only be in any one of the statonary states. Therefore, the energy of a free particle confined to a line of length L can be only one of the following values:

$$E_1 = \frac{h^2}{8mL^2}, E_2 = \frac{4h^2}{8mL^2}$$

$$E_3 = \frac{9h^2}{8mL^2}, E_4 = \frac{16h^2}{8mL^2} \ldots$$

and not *any* real number between zero and infinity as expected from classical mechanics. The integers n specifying the quantitized energy state of a system are called *quantum numbers*.

The stationary states of a quantum system are usually shown through energy level diagrams. In Fig. 13.8 the first five stationary states of a particle moving freely along a line of length L have been shown. The stationary state of the lowest energy in the energy-level diagram is called the *ground state*. The stationary state with energy $E_1 = h^2/(8mL^2)$ is the ground state of the quantum system under discussion.

$$E_n = \frac{n^2h^2}{8mL^2}, n = 1, 2, 3,$$

This example has brought out how a quantum concept when coupled with classical physics introduces discretisation in the allowed energy levels. We shall make use of

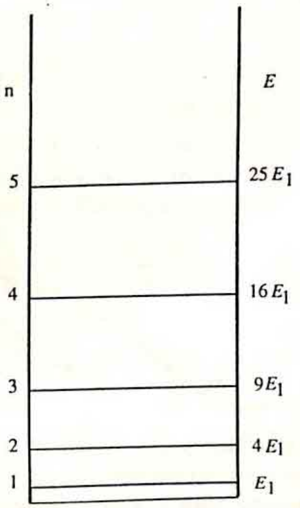


Figure 13.8: Energy level diagram showing the first five stationary states of a particle of mass m moving freely along a line of length L.

a similar argument and Rutherford's model of the hydrogen atom to calculate the energies of the special orbits corresponding to the stationary states. (This was not the argument used by Bohr, who simply proposed that those circular orbits for which the angular momentum is $(nh/2\pi)$ are stable).

Example 13.2: Find the energies in units of eV for the ground state and the first excited state for an electron confined to a line of length 10^{-10}m and also find the energies in units of MeV for the ground state and the first excited states of a neutron confined to a line of length 10^{-14}m.

Answer: The energies of the ground state and the first excited state of a free particle confined to a line correspond to $n = 1$ and $n = 2$ respectively in the formula

$$E_n = \frac{n^2h^2}{8mL^2}$$

$$E_1 = \frac{h^2}{8mL^2}, E_2 = \frac{4h^2}{8mL^2} = 4E_1.$$

For the first part of the problem we use the following data:

$$\begin{aligned} m_e &= 9.1 \times 10^{-31}\text{kg} \\ L &= 10^{-10}\text{m} \\ h &= 6.6 \times 10^{-34}\text{Js} \end{aligned}$$

By substituting these values we shall get the energy in joules.

$$\begin{aligned} E_1 &= \frac{(6.6 \times 10^{-34})^2}{8 \times 9.1 \times 10^{-31} \times (10^{-10})^2}\text{J} \\ &= 5.98 \times 10^{-18}\text{ J} \\ 1\text{eV} &= 1.6 \times 10^{-19}\text{J} \\ E_1 &= 5.98 \times 10^{-18} \\ &\times \frac{1}{1.6 \times 10^{-19}}\text{ eV } = 37\text{eV} \\ \text{and } E_2 &= 148\text{eV}. \end{aligned}$$

For the second part of the problem we use

$$\begin{aligned} m_n &= 1.67 \times 10^{-27}\text{kg} \\ L &= 10^{-14}\text{m} \\ E_1 &= \frac{(6.6 \times 10^{-34})^2}{8 \times 1.67 \times 10^{-27} \times (10^{-14})^2}\text{J} \\ &= 3.26 \times 10^{-13}\text{J} \\ 1\text{MeV} &= 1.6 \times 10^{-13}\text{J} \\ E_1 &= 2.0\text{MeV} \\ \text{and } E_2 &= 8.0\text{MeV}. \end{aligned}$$

13.4 Bohr model and the hydrogen spectrum

In Chapter 8 of the class XI textbook you learnt that the planets, because they are bound to the Sun by an attractive inverse square law force, move in elliptic orbits. The electronic orbits in a hydrogen atom will, therefore, also be elliptic. We make an additional assumption that stationary states of a hydrogen atom correspond to circular orbits. This assummption is not essential but

miraculously does not come in the way of obtaining the exact quantum mechanical result for the allowed energies of the hydrogen atom. The calculation of energies for stationary states corresponding to the circular orbits is as simple as finding the quantum energies of a free particle confined to a line of fixed length.

For an electron moving with a uniform speed in a circular orbit of a given radius, the centripetal acceleration is provided by the inverse square Coloumb force of attraction between the electron and the nucleus. The nucleus of the hydrogen atom and the electron each carry one unit of fundamental charge, e, but of opposite signs. It can be easily checked that the gravitational attraction between the electron and the proton masses is weaker than the Coulombic attraction by a factor of 10^{-40} and its contribution to the centripetal force can be neglected.

Let m be the mass of the electron. As the nucleus is about 2000 times heavier than m, it is reasonable to assume that it remains at rest. The centripetal acceleration for a particle moving in a circular orbit of radium r with speed v is v^2/r. Therefore, Newton's second law of motion gives

$$\frac{mv^2}{r} = \frac{e^2}{4\pi\epsilon_0 r^2}, \tag{13.6}$$

giving

$$r = \frac{e^2}{4\pi\epsilon_0 mv^2}. \tag{13.7}$$

A circular orbit is taken to be a stationary state *if its circumference contains integral numbers of de Broglie wavelengths.* This is graphically shown in Fig. 13.9 and can be mathematically expressed as

$$2\pi r = n\lambda = n\left(\frac{h}{mv}\right) \tag{13.8}$$

where $n = 1, 2, 3, \ldots$

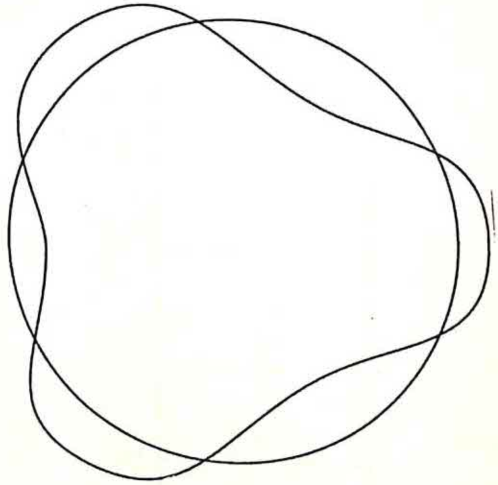


Figure 13.9: The orbits satisfying the conditions of stationary states with three waves. The radius of the orbit with n waves can be estimated from Eq. (13.11b).

This relation can also be read as a quatisation condition on the angular momentum, L, of the electron in the hydrogen atom,

$$L \equiv mvr = \frac{nh}{2\pi},$$
$$\text{where} \quad n = 1, 2, 3, 4, 5, \ldots \tag{13.9}$$

This is the famous Bohr quantisation condition. He had introduced it in physics as an adhoc postulate in violation of the known ideas of classical physics. It restricts the stationary states to those circular orbits in which the angular momentum is an integral multiple of $(h/2\pi)$. An additional postulate is that in violation of electrodynamics, accelerated motion of electronic charge in these orbits is not accompanied by emission of electromagnetic radiation. By this act of genius, Bohr 'provided' stability to the hydrogen atom! You may have already noticed that the dimensions of the Planck constant, h, are $(ML^2/T^2) \times T$. These are the dimensions of angular momentum. This perhaps had been in Bohr's mind when he made the intelligent guess.

Equation (13.8) gives

$$v = \frac{nh}{2\pi mr} \tag{13.10}$$

Equations (13.7) and (13.10) can be easily solved for v and r. The speed of the electron and the radius of the orbit corresponding to the stationary state marked by the label n are given by the expressions:

$$v = \frac{1}{n}\left(\frac{e^2}{4\pi\epsilon_0}\right)\frac{1}{(h/2\pi)} \tag{13.11a}$$

and

$$r = \frac{n^2}{m}\left(\frac{h}{2\pi}\right)^2\left(\frac{4\pi\epsilon_0}{e^2}\right) \tag{13.11b}$$

You know that the speed of light in vacuum, c, is a universal constant and its value is 3×10^8 m/s. To estimate how close the speed of electron in hydrogen atom is to the speed of light in vacuum, we multiply the numerator and denominator of Eq. (13.11a) with c:

It gives

$$v = \frac{c}{n}\left(\frac{e^2}{4\pi\epsilon_0(h/2\pi)c}\right) \tag{13.12}$$

The parenthesis of the above expression gives a dimensionless constant, which is called the *fine structure constant*, α. It occurs naturally in quantum processes involving electromagnetic forces. The numerical value of α can be easily computed by substituting for c, h and e. The fine structure constant

$$\alpha \equiv \frac{e^2}{4\pi\epsilon_0(h/2\pi)c} = \frac{1}{137} \tag{13.13}$$

Substituting this value in Eq. (13.12), we get

$$v = \frac{1}{137}\,\frac{c}{n} \tag{13.14}$$

From this equation we see that the speed of the electron in the innermost orbit, the stationary state with $n = 1$, is 1/137 of the speed of light in vacuum. The orbital speed in the outer orbits falls by the factor of n. The size of the innermost orbit is called the Bohr radius, a_0. Its value can be computed from Eq. (13.11b). It gives $a_0 = 5.29 \times 10^{-11}$m. It can also be seen from Eq. (13.11b) that the radii of the orbits increase in ratio of n^2. The parameter n is called the *principal quantum number*.

The energy of the electron in the stationary states of the hydrogen atom can be obtained by adding its kinetic energy and the potential energy. The kinetic energy is $(1/2)mv^2$ and the potential energy is $[(-e^2)/(4\pi\epsilon_0 r)]$. The total energy of the electron, E, gives the internal energy of the hydrogen atom,

$$E = \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r}. \tag{13.15}$$

We substitute the expressions for v and r. It gives the energies of the stationary states of the hydrogen atom,

$$E_n = \frac{1}{2}\,\frac{mc^2}{n^2}\left(\frac{e^2}{4\pi\epsilon_0(h/2\pi)c}\right)^2 - \frac{e^2}{4\pi\epsilon_0}\,\frac{me^2c^2}{4\pi\epsilon_0 n^2(h/2\pi)^2c^2}$$

or

$$\begin{aligned} E_n &= -\frac{1}{2}\,\frac{mc^2}{n^2}\left(\frac{e^2}{4\pi\epsilon_0(h/2\pi)c}\right)^2 \\ &= -\frac{1}{2}\,\frac{mc^2}{n^2}\alpha^2. \end{aligned} \tag{13.16}$$

The mass of the electron is 9.1×10^{-31} kg. We express $(1/2)mc^2\alpha^2$ in eV. One eV is the energy gained by an electron when it is accelerated through an electric field with a potential difference of 1 V.

$$\begin{aligned} &\frac{1}{2}mc^2\alpha^2 \\ &= \frac{1}{2} \times \frac{9.1 \times 10^{-31} \times (3.0 \times 10^8)^2}{1.602 \times 10^{-19}} \\ &\times \left(\frac{1}{137}\right)^2 \text{ eV}. \end{aligned}$$

$$= 13.6\text{eV}. \qquad (13.17)$$

Substituting the value of $(1/2)mc^2\alpha^2$ in Eq.(13.16) the energy of the hydrogen atom in its stationary states can be expressed in units of eV,

$$E_n = \frac{-13.6}{n^2}\text{eV}, \cdots \qquad (13.18)$$

The innermost orbit has the lowest energy equal to -13.6eV. It is, therefore, called the ground state. The energy of the other states are

$$\frac{-13.6}{4}\text{eV}, \ \frac{-13.6}{9}\text{eV}, \ \frac{-13.6}{16}\text{eV}$$

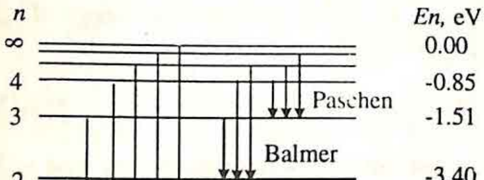


Figure 13.10: Energy level diagram for the hydrogen atom. The stationary state with $n = 1$ is the state of lowest energy and is called the ground state. A few of the transitions in the Lyman, Balmer and Paschen series are also shown.

Normally the hydrogen atom will be in its ground state. It is nearest to the nucleus and has a size given by the Bohr radius. When the hydrogen atom receives energy by processes such as electron collisions, the electron can make a transition to states with higher energy, which are therefore called the excited states. You may have noticed that the energy of the hydrogen atom in the stationary state is negative. It means that the electron is bound. Energy will be required to remove an electron from the hydrogen atom to a distance infinitely far away from its nucleus. The minimum energy required to free the electron from the ground state of the hydrogen atom is 13.6 eV. It is called the *ionisation energy*.

The energy level diagram for the stationary states of hydrogen atom is given in Fig. 13.10. The principal quantum number n labels the stationary states in ascending order of energy.

It was shown by Sommerfeld that when the restriction of circular orbits is relaxed, Eq. (13.16), continues to hold even for elliptic orbits.

Bohr had postulated that when a hydrogen atom makes a transition from the state with quantum number n_i to the state with quantum number n_f $(n_f < n_i)$, the difference of energy is carried away by a photon of frequency ν_{if} such that

$$h\nu_{if} = E_{n_i} - E_{n_f}. \qquad (13.19)$$

Using Eq. (13.16), for E_f and E_i, we get

$$h\nu_{if} = \frac{1}{2}\, mc^2\alpha^2 \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \qquad (13.20)$$

This is the Rydberg formula for the spectrum of the hydrogen atom. The spectral lines corresponding to $n_f = 2$ and $n_i = 3, 4, 5\ldots$ are called the Balmer series and lie in the visible part of the hydrogen spectrum. The spectral lines corresponding to transitions from $n_i = 2, 3, 4\ldots$ to $n_f = 1$ lie in the ultraviolet region and correspond to the

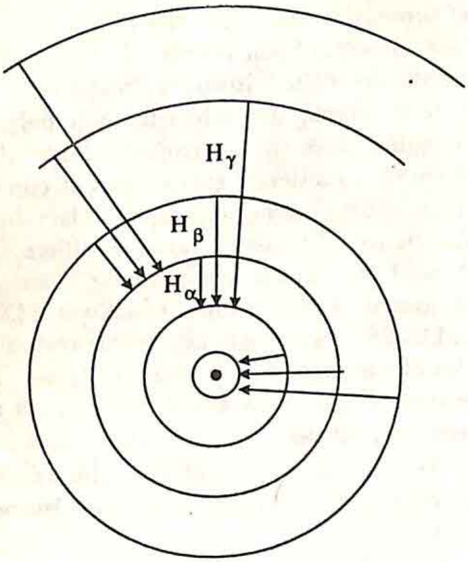


Figure 13.11: Diagram shows the circular orbits for the first few stationary states of the hydrogen atom. Each group of arrows represents a group of energy transitions which give rise to a spectral series. H_α, H_β, and H_γ are the spectral lines in the visible region and correspond to Balmer series.

Lyman series. Transitions for some prominent spectral lines of hydrogen are shown in Fig. (13.11). The impressive numerical agreement between the theoretically calculated values of the spectral frequencies of the hydrogen atom and the experimentally observed spectrum is one of the most beautiful success stories of physics. Niels Bohr was awarded the 1922 Nobel prize for physics for this work.

The spectrum of the hydrogen atom was determined first in the late nineteenth century. At that time, the few lines in the visible region were noted by Balmer, a school teacher in Switzerland, to have inverse wavelengths given by Eq. (13.20) with $n_f = 2$ and $n_i = 3, 4, 5, \ldots$. Later when spectra in the ultraviolet region were analyzed by the U.S. physicist Lyman, he found that the same formula fit the data, but with $n_i = 2, 3, 4, \ldots \infty$ and $n_f = 1$. Similar observations were made by other spectroscopists (Paschen, Pfund) in the infrared region. These were all shown by Rydberg to orginate from a single simple formula, namely Eq. (13.20). Since the hydrogen atom is the simplest of all atoms, with just *one* electron, and it has such a simple, characteristic spectrum, its understanding is of fundamental importance. Hence the significance of Bohr's work.

Example 13.3: Using the Rydberg formula, calculate the wavelengths of the first four spectral lines in the Balmer series of the hydrogen spectrum.

Answer: The Rydberg formula is

$$\frac{hc}{\lambda_{if}} = \frac{1}{2}mc^2\alpha^2\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

The wavelengths of the first four lines in the Balmer series correspond to transitions from $n_i = 3, 4, 5, 6$ to $n_f = 2$. We know $(1/2)mc^2\alpha^2 = 13.6\text{eV} = 21.76 \times 10^{-19}\text{J}$. Therefore,

$$\begin{aligned}\lambda_{i_2} &= \frac{hc}{21.76 \times 10^{-19}\left(\frac{1}{4} - \frac{1}{n_i^2}\right)}\,\text{m} \\ &= \frac{6.625 \times 10^{-34} \times 3 \times 10^8 \times 4n_i^2}{21.76 \times 10^{-19} \times (n_i^2 - 4)}\,\text{m} \\ &= \frac{3.653n_i^2}{(n_i^2-4)} \times 10^{-7}\,\text{m} = \frac{3653n_i^2}{(n_i^2-4)}\,\text{Å}\end{aligned}$$

Substituting $n_i = 3, 4, 5$ and 6, we get $\lambda_{32} = 6575$ Å, $\lambda_{42} = 4870$ Å, $\lambda_{52} = 4348$ Å and $\lambda_{62} = 4109$ Å.

The physics of atoms with two electrons or more is more complex than that of the hydrogen atom and cannot be tackled with

the rules used by Bohr. We will not pursue further the arrangement of electrons and the energy states of atoms more complex than hydrogen (e.g. He) and shift our attention to the study of the structure of the nucleus. We begin by a discussion of nuclear (or atomic) masses, which leads us to the fact that there are two kinds of fundamental particles namely, protons and neutons, that constitute all nuclei.

13.5 Atomic masses

The mass of the carbon atom, ^{12}C, is 1.992678×10^{-26}kg. It is indeed very small. Therefore, it is convenient to introduce a mass unit for expressing the mass of atoms. This unit is now defined by taking mass of ^{12}C atom to be 12 atomic mass unit(u). With this definition

$$\begin{aligned} 1\text{u} &= \frac{\text{mass } ^{12}\text{C atom}}{12} \\ &= \frac{1.992678 \times 10^{-26}}{12}\text{kg} \\ \text{or } 1\text{u} &= 1.660565 \times 10^{-27}\text{kg}. \qquad (13.21) \end{aligned}$$

In Table 13.3, given at the end of this Chapter, the atomic masses of various elements expressed in u have been listed. It may be noted that masses of many atoms are close to being integral multiples of the mass of an hydrogen atom. There are however, many striking exceptions to this rule. For example, the atomic mass of the chlorine atom is 35.46 u.

The mass spectrometer is an instrument which is an improved version of Thomson's apparatus for the measurement of e/M. The measurement of e/M of an atom after being ionised gives it mass M, because the value of the electric charge e is known. The measurement of atomic masses revealed the existence of different varieties of atoms of the same element. The different types of atoms of the same element show similar chemical properties but differ from each other in their masses. Such atoms of the same elements are called isotopes (meaning same place in Greek; this refers to their being in the same place in the periodic table of elements). Practically every element consists of a mixture of several isotopes. The relative abundance of different isotopes differs from element to element. For example, chlorine is composed of two isotopes of masses 34.98 u and 36.98 u, which are nearly integral multiples of the mass of a hydrogen atom. Their relative abundances are 75.4 and 24.6 percent, respectively. You can verify that mass of the natural chlorine atom is the weighted average of the masses of the two isotopes, that is,

$$\begin{aligned} &\text{mass of natural chlorine atom} \\ &= \frac{75.4 \times 34.98 + 24.6 \times 36.98}{100}\text{u} \\ &= 35.47\text{u} \qquad (13.22) \end{aligned}$$

Even the hydrogen atom has three isotopes with masses of 1.0078 u, 2.0141 u and 3.0160 u. The nucleus of the lightest atom of hydrogen which has a relative abundance of 99.985 percent is called the proton. The mass of the proton is

$$\begin{aligned} m_p &= 1.0073\,\text{u} \\ &= 1.6726 \times 10^{-27}\text{kg}. \qquad (13.23) \end{aligned}$$

This is equal to the mass of the hydrogen atom, which is 1.0078 u, minus the mass of the single electron, $m_e = 0.00055$ u, it contains. The other two isotopes of hydrogen are called deuterium and tritium. Tritium nuclei being unstable do not occur naturally and are produced in laboratory by processes involving nuclear changes.

The proton carries one unit of fundamental charge and is stable. The positive charge in the nucleus is that of the protons. If we assume that the nucleus does not contain any

negative charge, the total number of protons in the nucleus of an atom has to be exactly equal to the total number of electrons it contains.

The nuclei of deuterium and tritium, as they are isotopes of hydrogen, contain only one proton each. But the ratios of the masses of nuclei of hydrogen, deuterium and tritium are as 1:2:3. Therefore, the nuclei of deuterium and tritium must contain in addition to a proton, some neutral matter. The amount of neutral matter present in the nuclei of deuterium and tritium measured in the unit of mass of a proton are approximately equal to one and two, respectively. It indicates that nuclei of atoms, in addition to protons, contain neutral matter in multiples of a basic unit. This hypothesis was verified in 1932 by James Chadwick who observed emission of neutral radiation when beryllium nuclei were bombarded with alpha particles. It was found that this neutral radiation could knock out protons from light nuclei such as those of helium, carbon and nitrogen. Conservation of energy and momentum show that if this neutral radiation consisted of photons the energy of the photons will be much higher than is available in the bombardment of beryllium nuclei with α-particles. The clue to answering this riddle, which Chadwick solved, is to assume that the neutral radiation consists of neutral particles called neutrons.

Chadwick estimated the mass of a neutron by comparing the maximum velocity, U_p, imparted to a hydrogen atom when a neutron collides it head-on with the maximum velocity, U_n, imparted to a nitrogen atom when a neutron collides it head on. This problem is identical to the calculation of the speed U with which a stationary target of mass M will move forward on being hit head-on by a particle of mass m moving with speed V. This is an elementary problem of elastic collison in mechanics and can be solved by applying the principles of conservation of momentum and energy. Let V' be the speed of the projectile after the head-on collision. The conservation of momentum gives the equation

$$mV = MU + mV' \tag{13.24}$$

The conservation of energy gives the equation

$$\frac{1}{2}mV^2 = \frac{1}{2}MU^2 + \frac{1}{2}mV'^2 \tag{13.25}$$

Elimination of V' from Eq.(13.24) and (13.25) gives

$$U = \frac{2mV}{m+M}. \tag{13.26}$$

The speed U_p and U_n can now be obtained from Eq. (13.26) substituting $M = 1$ u for proton and M = 14 u for neutron, respectively.

$$U_p = \frac{2mV}{m+1},$$
$$U_n = \frac{2mV}{m+14}.$$

Then,

$$\frac{m+14}{m+1} = \frac{U_p}{U_n}. \tag{13.27}$$

Chadwick measured U_p and U_n. The approximate values of these velocities as measured by him were $U_p = 3.7 \times 10^9$ cm/s, $U_n = 4.7 \times 10^8$ cm/s. By substituting these values in Eq. (13.27) Chadwick estimated $m \simeq 0.9$ u. The mass of neutron is now known with high accuracy. It is

$$m_n = 1.00866\,\text{u} = 1.6749 \times 10^{-27}\text{kg}.$$

Chadwick was awarded the 1935 Nobel prize for physics for his discovery of the neutron. A free neutron, unlike a free proton, is unstable. It decays into a proton, an electron and a neutrino (another elementary particle of matter) with a mean life of about 1000s. It is stable, however, inside the nucleus.

What force keeps the neutrons and protons confined to a tiny region? Protons which are positively charged do not like to stay together because of coulombic repulsion. The force which provides the binding of neutrons and protons inside the nucleus is neither electrical nor gravitational. It is a new type of force called the strong *nuclear force.* It does not depend on charge. It acts between a pair of neutrons which are electrically neutral. It also acts between a neutron proton pair and between a pair of protons with equal strength. Unlike gravitational and electrical forces the nuclear force is of short range, of order 2-3 fm. It is attractive but becomes strongly repulsive when the separation is less than about 1 fm (one fm = 10^{-15}m).

The graph of the potential energy of a pair of nuclear particles as a function of their separation is roughly shown in Fig. 13.12.

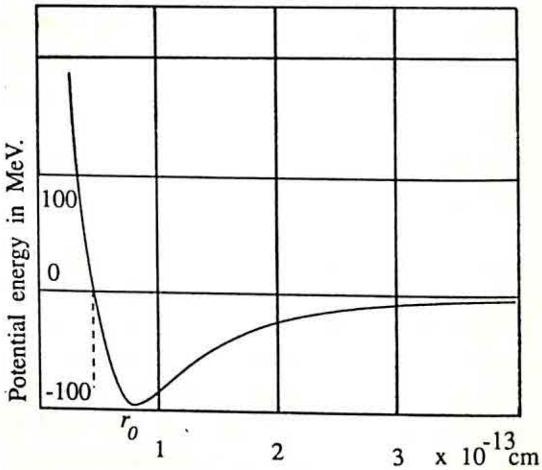


Figure 13.12: Potential energy of a pair of nucleous as a function of their separation. For a separation greater than r_0 the force is attractive and for separation less than r_0 the force is strongly repulsive. The attractive force is strongest when the separation is about 1 fm.

Neutrons and protons are almost identical in the sense that their masses are nearly the same and the nuclear force does not distinguish them. Therefore, the neutron and the proton have a common name, the *nucleon.* As the proton is positively charged and the neutron is electrically neutral the electromagnetic force can distinguish them.

The composition of a nucleus can now be described in terms of the following terms and symbols:

Z = atomic number = number of protons
N = neutron number = number of neutrons
$A = Z + N$ = mass number = total number of nucleons (total number of neutrons and protons).

Nuclear species are described according to the notation

^A_ZX

where X is the chemical symbol of the species For example, the nucleus of gold is denoted by $^{197}_{79}$Au. It contains 197 nucleons, of which 79 are protons and 118 neutrons.

The composition of isotopes of an element can now be readily explained. The nuclei of isotopes of a given element contain the same number of protons but differ from each other in their number of neutrons. Deuterium, $^2_1$H which is an isotope of hydrogen contains one proton and one neutron. Its other isotope tritium, $^3_1$H contains one proton and two neutrons. We have already mentioned that chemical properties of elements depend on their electronic structure. As the atoms of isotopes have identical electronic structure they have identical chemical behaviour and are placed in the same location in the periodic table.

13.6 Binding energy

Using the ideas of the special theory of relatively Einstein had derived a formula giving the equivalence of mass and energy. The famous mass-energy relation is

$$E = mc^2 \tag{13.28}$$

In this formula c is the speed of light in vacuum (3×10^8 m/s), m is the mass of the object and E is its energy. It states that an object of mass m has an energy content equal to its product with the square of the speed of light. An object of mass 1 kg has an energy content of 9×10^{16} joules ! Equivalently the mass associated with an amount of energy E is equal to E/c^2. Einstein's mass energy relation raised the hope of liberating energy contained in the mass of ordinary matter. Because a proton lives for more than 10^{32} years and free neutrons decay into protons and electrons, the total number of nucleons (neutrons + protons) remains unchanged in all processes. The 9×10^{16}J of energy contained in the 1 kg of matter cannot thus be liberated because nature forbids the disappearance of nucleons unless by annihilation with antinucleons. But in our part of the universe, matter is made out of nucleons and electrons. Anti-matter which can annihilate matter is found in cosmic rays and can also be created in high energy laboratories.

You may like to note that in units of MeV (million electron volt) the energy equivalent to mass of an electron $= m_e c^2 = 0.511$ MeV. Similarly one has energy equivalent to mass of proton $= m_p c^2 = 938.279$ MeV; energy equivalent to mass of a neutron $= m_n c^2 =$ 939.573 MeV, and

$$1 \text{ u} = 931.501 \text{ MeV}/c^2$$

However, it is found that the mass of a stable nucleus is less than the total mass of its nucleons. It means that if a certain number of neutrons and protons are brought together to form a nucleus of a certain charge and mass, energy will be released in the process. This mass difference can be given in terms of energy by Einstein's mass energy relation. For example, in the case of oxygen, we have:

mass of 8 neutrons = 8×1.00865 u
mass of 8 protons = 8×1.00727 u
mass of 8 electrons = 8×1.00055 u
Total mass = 16.13176 u

which can be compared with the atomic weight of oxygen $^{16}_{8}$O, which is 16.0000 u. The difference in mass of a nucleus and its constituents is called the *mass defect*. The energy equivalent to the mass defect of a nucleus is called its *binding energy*. The expression for the binding energy per nucleon (binding fraction) of the nucleus $^{A}_{Z}X$ of the atom X of mass m is

$$\begin{aligned} \overline{B} &\equiv \frac{BE}{A} \\ &= \frac{(Zm_p + (A-Z)m_n - m)c^2}{A} \end{aligned} \tag{13.29}$$

★ Antinucleons (e.g. antiprotons) and anti-electrons (called positrons) have been produced in the laboratory. Some of them have also been seen in natural processes. Antiparticles have the same mass as particles, but opposite charge and magnetic moment. ★

In Fig. 13.13 $\overline{B}$ has been plotted for all the known nuclei. We notice the following:

(i) The binding energy per nucleon, $\overline{B}$ is approximately equal to about 8 MeV for nuclei of middle weight (or nucleon number).

(ii) $\overline{B}$ is less than 8 MeV (per nucleon) for both very light nuclei ($Z \leq 10$) and very heavy nuclei ($Z \geq 70$).

13.7 Size of the nucleus

The Geiger-Marsden experiments on the scattering of alpha particles from atoms of gold revealed that the size of the gold nucleus has to be less than 4×10^{-14}m, which is the distance of the closest approach of α - particles of energy 5.5MeV. For very high

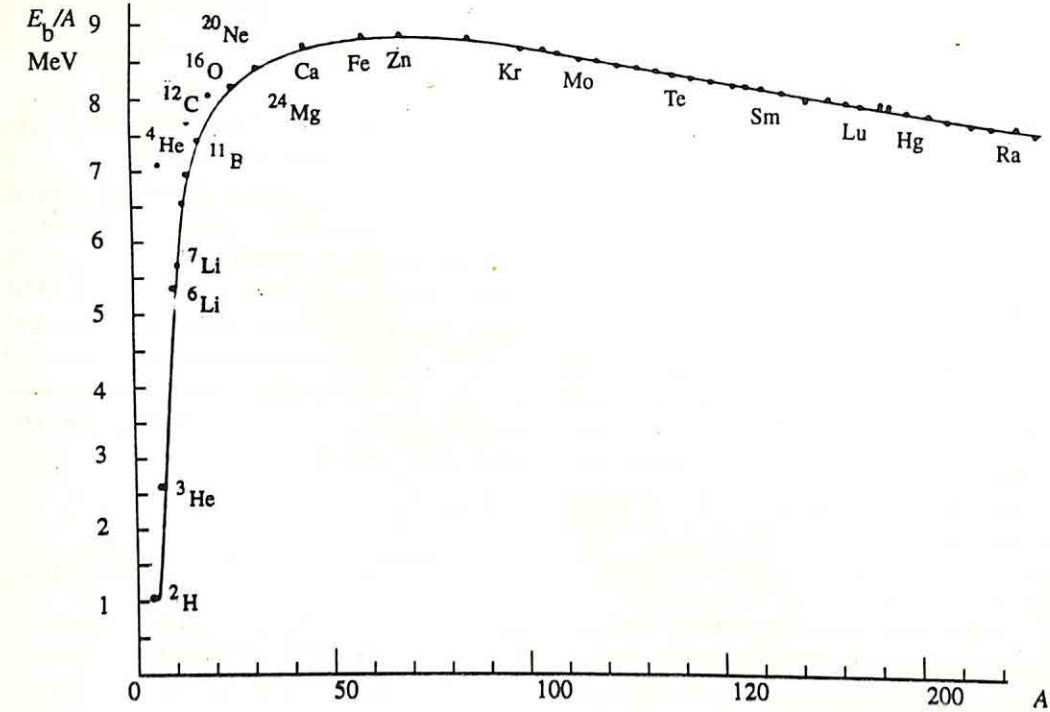


Figure 13.13: Plot of binding fraction, $\bar{B}$ or binding energy per nucleon as a function of mass number, A.

energy alpha particles the nature of scattering is affected by short range nuclear forces, and will differ from Rutherford's calculations, which are based on a purely coulombic repulsion between the gold nucleus and the alpha-particle. By noting when such deviations occur, nuclear sizes can be inferred.

By performing scattering experiments with fast electrons, sizes of nuclei of different elements have been accurately measured. Assuming nuclei to be spherical, their volumes can be estimated. These measurements have revealed that the volume of a nucleus is directly proportional to its mass number A. In other words the density of nuclear matter is independent of the size of the nucleus. Nuclear matter therefore behaves like a liquid of constant density. Different nuclei are like drops of this liquid, of different sizes. As the volume of sphere is proportional to the cube of its radius, the following empirical relation has been found to hold between the radius R of nucleus and its mass number A:

$$R = R_0 A^{1/3} \tag{13.30}$$

$$\text{where } R_0 = 1.1 \times 10^{-15}\,\text{m} \tag{13.31}$$

which is of the order of the range of nuclear force. From this formula we can compute the density of nuclear matter. For example, the density of matter in the nucleus of iron can be estimated from the data,

$$\begin{aligned} m_{Fe} &= 55.85\,\text{u} = 9.27 \times 10^{-26}\text{kg} \\ A &= 56 \\ \rho_{nuclear} &= \frac{9.27 \times 10^{-26}}{(1.1 \times 10^{-15})^3} \\ &\quad \times \frac{1}{56 \times (4\pi/3)}\ \text{kgm}^{-3} \\ &= 2.9 \times 10^{17}\,\text{kgm}^{-3} \qquad (13.32) \end{aligned}$$

It is very large when compared to the density of ordinary matter, say water, which is 10^3kgm^{-3}. In Chapter 15 which deals with the universe you will come across stars which have the density of the nuclear matter and are called neutron stars.

Neutrons and protons in the nucleus are held inside by a complex action of forces. Nucleons attract each other with the nuclear force. In addition, protons being positively charged exert coulombic repulsion. This is in general a small effect, except for very heavy nuclei ($Z \geq 70$). Like an atom which exists only in electronic configurations corresponding to stationary states, the nucleus exists in nucleonic configurations which correspond to nuclear stationary states. The stationary state of the lowest energy is called the ground state.

The nucleus like an atom can be excited from its ground state to stationary states of higher energy. This occurs in processes which impart energy to the nucleus. An excited nucleus like an excited atom can make a transition to its ground state by emitting the difference of energy as a photon. Photons emitted in nuclear transitions have energies of the order of MeV. Their wavelength is shorter than that of X-rays. These radiations were discovered by Becquerel and are called gamma-rays (γ-rays).

Unlike an atom a nucleus can be unstable even in its ground state. It then tries to reach states of greater stability by emission of an alpha-particle ($^4_2$He) or a beta-particle (electron). Invariably the α- emission or the β-emission leave the daughter nucleus in an excited state which in turn emits γ-photons for reaching a stable state.

Radioactivity is the generic name for these processes. It was discovered by A.H. Becquerel in 1896 purely by accident. Radioactivity has played a key role in understanding the structure of the atom and its nucleus. We study it next.

13.8 Radioactivity

The broad features of the three different types of radioactivity (α, β and γ emission) are as follows:

13.8.1 Alpha decay

Alpha decay is a process involving the emission of a fast-moving helium nucleus (an alpha-particle) by nuclei which generally contain 210 or more nucleons. Since $^4_2$He contains two protons and two neutrons, after an alpha emission the parent nucleus is transformed into a daughter nucleus which has an atomic number smaller by two and atomic weight smaller by four. Transformation of the $^A_Z X$ nucleus into the $^{A-4}_{Z-2}Y$ nucleus by an alpha decay can be expressed by the equation

$$^A_Z X \rightarrow ^{A-4}_{Z-2} Y + ^4_2 \text{He}. \qquad (13.33)$$

The energy, Q, released in this process can be obtained from Einstein's mass energy relation.
It is given by the expression

$$Q = (m_X - m_Y - m_{\text{He}})c^2. \qquad (13.34)$$

This energy is shared by the daughter nucleus, $^{A-4}_{Z-2}Y$ and the alpha-particle, $^4_2$He.

As the parent nucleus $^A_Z X$ is at rest before it undergoes alpha- decay, alpha particles are emitted with fixed energy, which can be calculated by applying the principles of conservation of energy and momentum. Let v_{He} and v_Y be the velocities of the alpha-particle and the daughter nucleus, $^{A-4}_{Z-2}Y$. The principle of conservation of momentum gives

$$m_Y \, v_Y = m_{\text{He}} \, v_{\text{He}}. \qquad (13.35)$$

By equating the sum of kinetic energies of the nucleus Y and the alpha particle to the energy released in the alpha-decay, we have another equation

$$\frac{1}{2}m_{\text{He}}v_{\text{He}}^2 + \frac{1}{2}m_Y v_Y^2 = Q. \qquad (13.36)$$

By substituting for v_Y from Eq. (13.35) in Eq. (13.36) you can easily obtain

$$\frac{1}{2}m_{\text{He}}v_{\text{He}}^2 = \frac{m_Y}{m_Y + m_{\text{He}}}\,Q. \qquad (13.37)$$

If we substitute $m_Y \simeq (A-4)$ u and $m_{\text{He}} \simeq 4$ u in Eq.(13.37) the kinetic energy carried by the alpha-particle can be approximated by the relation

$$KE_{\text{He}} \equiv \frac{1}{2}m_{\text{He}}v_{\text{He}}^2 \simeq \frac{A-4}{A}\,Q. \qquad (13.38)$$

For example, in the decay of $^{222}_{86}$Rn, $Q = 5.58$ MeV and $KE_{\text{He}} = 5.486$ MeV. The velocity of the alpha-particle emitted by $^{222}_{86}$Rn can be easily estimated from its kinetic energy.

$$v_{\text{He}} = \sqrt{\frac{2\times 5.486\times 1.6\times 10^{-13}}{4.00\times 1.66\times 10^{-27}}}\ \text{m/s}$$
$$= 1.62\times 10^7\ \text{m/s}.$$

13.8.2 Beta decay

Beta (β) decay is another process in which the nucleus by emitting an electron achieves greater stability. A neutron inside the nucleus is transformed into a proton by emitting an electron. The atomic number of the resulting element is increased by one but the atomic weight does not change because of the negligibly small mass of the electron. In a beta- decay process by emitting an electron the nucleus ^A_ZX is transformed into the nucleus $^A_{Z+1}Y$. We therefore expect this process to be described by relation

$$^A_ZX \rightarrow {}_{Z+1}{}^AY + e^-. \qquad (13.39)$$

As the decay process given in Eq. (13.39) is similar in form to an alpha- decay, we also expect that in the beta-decay electrons will be emitted with fixed energy. When the kinetic energies of electrons emitted from a given radioactive nucleus were measured, they revealed that beta-decay electrons do not come out with fixed energy. It varies continuously from zero to a maximum value. To account for this behaviour Pauli postulated in 1930 that beta-decay is not a two-body decay process. An uncharged particle, which interacts very weakly with matter and hence escapes undetected, is emitted along with an electron. This particle was called antineutrino, $\bar{\nu}$. Instead of Eq. (13.39), beta-decay is correctly described by the relation

$$^A_ZX \rightarrow {}_{Z+1}{}^AY + e^- + \bar{\nu}. \qquad (13.40)$$

Also, it is seen that a free neutron decays into a proton, electron and antineutrino, i.e.

$$n \rightarrow p + e^- + \bar{\nu}. \qquad (13.41a)$$

There is another type of beta-decay in which a nucleus emits an antielectron, called positron, and a neutrino. Such decays are described by the relation

$$^A_ZX \rightarrow {}_{Z-1}{}^AY + e^+ + \nu \qquad (13.41b)$$

Neutrinos (antineutrinos) interact so weakly with matter that they even penetrate through earth without being absorbed. By ingenious experiments neutrinos have been detected and their physical properties such as mass and spin or intrinsic angular momentum have been measured. You may note that nucleon number is conserved in both α and β-emissions.

13.8.3 Gamma decay

Like an excited atom, an excited nucleus can make a transition to a state of lower energy by emitting a photon. We have seen that energies of the atomic states of hydrogen are of the order of electron volts. Therefore, the wavelength of light emitted in atomic transitions correspond to photons having energy of the order of electron volts. The energy of the nuclear states is of the order of million electron volts (MeV). Therefore, the photons emitted by nuclei can have energy of

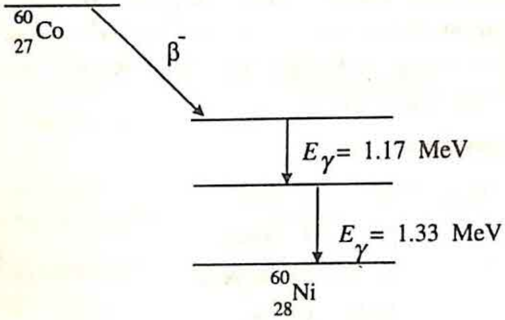


Figure 13.14: Energy-level diagram showing the emission of γ-rays by a $^{60}_{27}Co$ nucleus subsequent to beta decay.

the order of several MeV. The wavelength of photons of such energy is a fraction of an angstrom. The short wavelength electromagnetic waves emitted by nuclei are called the gamma rays.

Most radioisotopes after an alpha-decay or a beta-decay leave the daughter nucleus in an excited state. The daughter nucleus by single transition or sometimes by successive transitions reaches the ground state by emitting one or more gamma-rays. A well known example of such a process is the decay of $^{60}_{27}Co$. By beta-emission the $^{60}_{27}Co$ nucleus is first transformed into an excited $^{60}_{28}Ni$ nucleus, which in turn reaches the ground state by emitting photons of energies 1.17 MeV and 1.33 MeV. In Fig. 13.14 we have shown this process through an energy-level diagram.

13.8.4 Radioactive decay law

Radioactive emissions are characterised by a given probability per unit time for a nucleus to decay. Let this probability be λ. Although it is not possible to predict whether a particular nucleus will decay in a time interval, dt, the decay behaviour of a collection of a large number of nuclei can be predicted

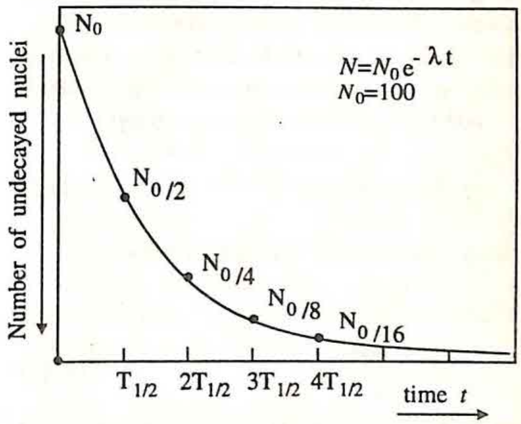


Figure 13.15: Exponential decay curve. After lapse of $T_{1/2}$ s population of given species drops by a factor of 2.

accurately (by using statistical arguments). Let $N(t)$ be the number of radioactive nuclei species present in a sample at time instant t. The number $|\ dN\ |$ of radioactive decays in time interval dt can be obtained from λ. The probability for a single nucleus to decay in the interval dt will be proportional to the time, i.e. equal to λ dt. As the number of radioactive nuclei present is N, dN will be given by the equation

$$|dN| = \lambda N(t)dt \qquad (13.42)$$

In calculus dN denotes the change in N. As dN is negative, Eq. (13.42) should be written as

$$dN = -\lambda N(t)dt \qquad (13.43)$$

$$\text{or} \quad \frac{dN}{dt} = -\lambda N(t). \qquad (13.44)$$

It is easily seen that

$$N(t) = N(0)e^{-\lambda t} \qquad (13.45)$$

is a solution of the differential equation given in Eq. (13.44). $N(0)$ is the number of species present at $t = 0$. Such a behaviour is called

an exponential decay. In Fig. 13.15 the exponential decay curve has been shown. Radioactive decays are characterised by a time constant $T_{1/2}$, called the *half-life*, which is related to λ. It is the time interval in which one-half the number of nuclei decay.

$$\frac{N(T_{1/2})}{N(0)} = \frac{1}{2} = e^{-\lambda T_{1/2}} \tag{13.46}$$

Taking the natural logarithm, we find

$$\begin{aligned} T_{1/2} &= \frac{\ln 2}{\lambda} \\ &= \frac{0.693}{\lambda} \end{aligned} \tag{13.47}$$

The half-life of radioactive nuclei can be as long as the estimated age of the universe, which is 10^{10} years, and can be shorter than 10^{-15}s. Those radioactive elements whose half-life is short compared to the age of the universe are not found in observable quantities in nature today. They have however been seen in experiments involving nuclear reactions. Tritium and plutonium belong to this category.

It is found useful to introduce a concept called the *rate of decay*, R. It is equal to the number of radioactive disintegrations in the sample taking place per second. Mathematically, it is represented as

$$R(t) = \left|\frac{dN}{dt}\right| = \lambda N(t) \tag{13.48}$$

As $N(t)$ varies exponentially so does $R(t)$. In SI units decay rate R is measured in terms of a unit appropriately called the curie (Ci) in honour of Madame Marie Sklodowska Curie (1867-1934). It is defined as

1 Ci (curie)	=	3.70×10^{10} disintegrations/s
1 mCi (milli curie)	=	3.70×10^{7} disintegrations/s
1 μCi (micro curie)	=	3.70×10^{4} disintegrations/s

Example 13.4: The half-life of $^{238}_{92}$U against alpha decay is 4.5×10^9 years. How many disintegrations per second occur in 1 g of $^{238}_{92}$U?

Answer:

$$\begin{aligned} T_{1/2} &= 4.5 \times 10^9 \text{ years} \\ &= 4.5 \times 10^9 \text{ years} \\ &\quad \times 3.16 \times 10^7 \text{ s/year} \\ &= 1.42 \times 10^{17} \text{ s} \end{aligned}$$

A kmol of an isotope has a mass equal to the atomic weight of that isotope expressed in kg. Hence 1 g of $^{238}_{92}$U contains

$$\frac{10^{-3}\text{kg}}{238\text{kg/kmol}} = 4.20 \times 10^{-6} \text{ k mol.}$$

One k mole of any isotope contains Avagadro's number of atoms, and so 1 g of $^{238}_{92}$U contains

$4.20 \times 10^{-6}\text{kmol} \times 6.025 \times 10^{26}$ atoms/ kmol $= 25.3 \times 10^{20}$ atoms.

The decay rate R is

$$\begin{aligned} R &= \lambda N \\ &= \frac{0.693}{T_{1/2}} \times N \\ &= \frac{0.693 \times 25.3 \times 10^{20}}{1.42 \times 10^{17}} \\ &= 1.23 \times 10^4 s^{-1} \end{aligned}$$

Example 13.5: Tritium ($^3_1$H) has a half-life of 12.5 years against beta decay. What fraction of a sample of pure tritium will remain undecayed after 25 years ?

Answer: By the definition of half life we can say that 1/2 of the initial tritium nuclei will remain undecayed after the first 12.5 years. In the next 12.5 years 1/2 of these nuclei will have decayed. Hence, the 1/4 of the

initial pure tritium will remain undecayed after 25 years.

Example 13.6: Determine the amount of $^{210}_{84}Po$ necessary to provide a source of alpha-particles of 5 mCi strength. The half-life of polonium is 138 days.

Answer:

$$R = N\lambda = \frac{0.693}{T_{1/2}} \times N$$

$$R = 5 \times 3.7 \times 10^7 \text{dis/s}$$

$$T_{1/2} = 138 \text{ days} = 138 \times 8.64 \times 10^4 \text{s} = 1.192 \times 10^7 \text{s}$$

$$N = \frac{5 \times 3.7 \times 10^7 \times 1.192 \times 10^7}{0.693} = 3.18 \times 10^{15} \text{ atoms}$$

Now 210 g of polonium contain 6.025×10^{23} atoms, so that the amount necessary to obtain a source of the required strength

$$= \frac{210 \times 3.18 \times 10^{15}}{6.025 \times 10^{23}} = 1.11 \times 10^{-6}\text{g}$$

13.9 Nuclear energy

The curve of average binding energy per nucleon, $\overline{B}$, give in Fig. 13.13 has a long flat region from about $A = 30$ to $A = 170$. In this region the binding energy per nucleon is nearly constant. However, for $A < 30$ and $A > 170$, $\overline{B}$ is less than the plateau-value. This feature of binding energy curve means that nuclei in the range $30 \leq A \leq 170$ are more tightly bound than nuclei with $A < 30$ and $A > 170$. Liberating energy by transmutation of a less tightly bound nucleus into more tightly bound nuclei through nuclear reactions provided the exciting possibility of releasing nuclear energy.

It should be fully appreciated that the chemical energy liberated or absorbed in chemical reaction is also a mass defect (binding energy) phenomenon. But the mass defects associated with chemical reactions of atoms are million times smaller than the mass defects involved in nuclear reactions.

Nuclear reactions, which can be practical sources of energy are of two broad types (a) fission reaction, and (b) fusion reaction.

13.10 Fission reaction

It is observed that a heavy nucleus($A > 230$) when excited splits into two lighter nuclei. The total mass of the lighter nuclei is less than the mass of the heavy nucleus. In 1938 O. Hahn and F. Strassmann found that barium, a medium weight element was one of the products when uranium was bombarded with neutrons. Nuclei of $^{235}_{92}U$ are sufficiently excited by the mere absorption of a neutron and split into two. Such a process is called *nuclear fission*. The amount of energy released in the splitting of a nucleus like that of $^{235}_{92}U$ is 0.9MeV/nucleon.

Most of the energy released in a fission reaction is in the form of the kinetic energy of the fission fragments. An example of a fission reaction is the splitting of a $^{235}_{92}U$ nucleus by capture of a slow neutron.

$$^{235}_{92}\text{U} + \text{n} \rightarrow \ ^{236}_{92}\text{U}$$

$$\rightarrow \ ^{144}_{56}\text{Ba} + ^{89}_{36}\text{Kr} + 3\text{n} + Q.$$

The Q value in this reaction is about 200 MeV.

13.11 Fusion reaction

By fusion of light nuclei into a heavy nucleus energy can also be liberated. In fact, fusion

of hydrogen nuclei into helium nucleus is the source of energy in all stars including our sun. One of the cycles which combines four protons to make an alpha particle is

$$^1\mathrm{H} + ^1\mathrm{H} \rightarrow ^2\mathrm{H} + e^+ + \nu + 0.41\mathrm{MeV} \quad \text{(a)}$$

$$^2\mathrm{H} + ^1\mathrm{H} \rightarrow ^3\mathrm{He} + \gamma + 5.51\mathrm{MeV} \quad \text{(b)}$$

$$^3\mathrm{He} + ^3\mathrm{He} \rightarrow ^4\mathrm{He} + 2\,^1\mathrm{H} + \gamma + 12.98\,\mathrm{MeV} \quad \text{(c)}$$

If we consider the combination $2(a)+2(b)+(c)$, the net effect is

$$4\,^1\mathrm{H} \rightarrow ^4\mathrm{He} + 2e^+ + 2\nu + 24.82\,\mathrm{MeV}.$$

In these reactions ^{1}H and ^{2}H denote hydrogen and deuteron, ^{4}He and ^{3}He denote helium and its isotope, e^+ denotes positron (antiparticle of electron) and γ denotes photon.

The fusion reactions take place under conditions of extreme temperature and pressure. This is necessary for the protons to have enough kinetic energy and thus overcome their mutual electrostatic repulsion so that they can come closer than the range of nuclear force. Such conditions exist in the interior of the sun, where the temperature is about 2×10^6K. Such extreme conditions cannot be easily arranged in a laboratory.

Although the possibility of using fusion reaction as a source of useful energy is being explored, the controlled liberation of nuclear energy through fission processes has been achieved in nuclear reactors. We describe it next.

13.12 Nuclear reactor

In 1939 Enrico Fermi (1901-1954) suggested that neutrons might be released in the fission of a uranium nucleus. If this were the case, and if the number of neutrons was more than one, some loss of neutrons could be tolerated and still there may be one neutron available for initiating another fission of uranium. Thus, a chain reaction, can be produced, which would continue to "burn" uranium. It was soon found that on an average $2\frac{1}{2}$ neutrons per fission of uranium nucleus are released. This number was kept secret during World War II.

But soon it was discovered that neutrons liberated were so energetic that they would escape instead of triggering another fission reaction. Also, it turns out that slow neutrons have a much higher intrinsic probability of inducing fission in $^{235}_{92}$U than fast neutrons.

The average energy of a neutron produced in fission of $^{235}_{92}$U is 2 MeV. These neutrons unless slowed down will escape from the reactor without interacting with the uranium nuclei, unless a very large amount of fissionable material is used for sustaining the chain reaction. What one needs to do is to slow down the fast neutrons by elastic scattering with light nuclei. We have seen in Chadwick's experiments that in an elastic collision with hydrogen the neutron almost comes to rest and protons carries away its energy. This is the same situation as when a marble hits head-on an identical marble at rest. Therefore in reactors light nuclei called *moderators* are provided along with the fissionable nuclei for slowing down fast neutrons. The moderators commonly used are water, heavy water (D_2O) and graphite. The Apsara reactor at the Bhabha Atomic Research Centre, Bombay, uses water as moderator. The other Indian reactors (Tarapur, Kalpakkam, Kota, Narora), which are used for power production, use heavy water as moderator.

Because of the use of moderator it is possible that the ratio, K, of number of fissions produced by a given generation of neutrons to the number of fission of the pre ceeding generation may be greater than one.

If this situation happens then the reaction rate and the reactor power increase exponentially. Unless the factor K is brought down very close to unity the reactor will become over critical and can even explode. The explosion of the Chernobyl Reactor in the Ukraine is a sad reminder that accidents in a nuclear reactor can be catastophic. The reaction rate is controlled through control-rods made out of neutron-absorbing material such as cadmium.

In addition to control-rods reactors are provided with safety-rods which when required can be inserted into the reactor and K can be reduced rapidly to less than unity.

As the energy released in nuclear reactions is a million times larger than in chemical reactions, nuclear power reactors need fuel million times less in weight than used in the thermal reactors of the same power capacity. However, in nuclear reactions highly radioactive elements are continuously produced.

The abundant ^{238}U isotope, which does not fission, on capturing a neutron leads to the formation of plutonium. The series of reactions involved is written as follows:

$$\begin{aligned} &^{238}\mathrm{U} + \mathrm{n} \rightarrow \ ^{239}\mathrm{U} \rightarrow \ ^{239}\mathrm{Np} + e^- + \overline{\nu} \\ &^{239}\mathrm{Np} \rightarrow \ ^{239}\mathrm{Pu} + e^- + \overline{\nu} \end{aligned} \tag{13.49}$$

where n is a neutron and e^- an electron. Plutonium which is an artificially produced element is highly radioactive and can also undergo fission under bombardment with slow neutrons. Unlike the waste of thermal power houses, which burn coal and leave ash, the waste of nuclear reactors is highly radioactive and extremely hazardous to all forms of life on the earth. It is, therefore, not surprising that nuclear reactors have become controversial.

13.13 Nuclear holocaust

In a single uranium fission about 0.9×235 MeV (≈ 200 MeV) of energy is liberated. If each nucleus of about 50 kg of ^{235}U undergoes fission the amount of energy involved is about 4×10^{15}J. This much amount of energy is equivalent to about 20,000 tons of TNT, enough for a superexplosion. Uncontrolled release of large nuclear energy is called an atomic explosion. On August 6, 1945 an atomic device was used in warfare for the first time. The US dropped an atom bomb on Hiroshima, Japan. The explosion was equivalent to 20,000 tons of TNT. Instantly the radioactive products devastated 10 sq km of the city which had 3,43,000 inhabitants. Of this number 66,000 were killed and 69,000 were injured; more than 67% of the city's structures were destroyed.

High temperature conditions for fusion reactions can be created by exploding a fission bomb. Super-explosions equivalent to 10 megatons of explosive power of TNT were tested in 1954. Such bombs which involve fusion of isotopes of hydrogen, deuterium and tritium are called hydrogen bombs. It is estimated that a nuclear arsenal sufficient to destroy every form of life on this planet several times over is in position to be triggered by the press of a button. Such a nuclear holocaust will not only destroy the life that exists now but its radioactive fallout will make this planet unfit for life for all times. Scenarios based on theoretical calculations predict a long *nuclear winter*, as the radioactive waste will hang like a cloud in the earth's atmosphere and will absorb the sun's radiation. Will man, the most intelligent form of life on this planet, use nuclear technology for generating energy to improve the quality of life or use it to destroy the planet? Only time can tell the answer to this question.

13.14 Molecules

You now know the structure of atoms. Atoms are electrically neutral systems made up of charged particles. In all atoms a negatively charged electron cloud surrounds the positively charged nucleus yet the net electric field at a distance of few atomic dimensions from the centre of the atom may not be traceable. However, when atoms come sufficiently close to each other, such that there is overlap of their electron clouds, they can interact with each other electrically. They may even prefer to stay together rather than to exist as separate neutral atoms. When this happens atoms are said to form molecules.

The phenomenon of molecule formation is extremely interesting and varied. It is responsible for many chemical and biological processes. We shall not discuss this in any detail here, but just outline a few physical ideas. A molecule is said to be stable if energy is required to split it into its constitutent atoms. The least energy requried to break a molecule is called its *dissociation energy*, E_d.

Atoms can combine in many ways to form molecules. Atoms of the same element can form stable molecules, for example, H_2, N_2, O_2, F_2. Atoms of different elements can also form stable diatomic molecules like carbon monoxide, CO_2, and even complex polyatomic molecules like sucrose, $C_{12}H_{22}O_{11}$. Biological molecules like DNA and RNA carry the blueprint of different life forms and exist in chains consisting of thousands of atoms.

Atoms join together by chemical bonding to form molecules. All chemical bonds form because electron clouds can be shared by two nuclei. The chemical bond in H_2 is formed because each of the two electron clouds is attracted to two protons simultaneously. As a rule, bonds form if energy is reduced when atoms come near each other.

Bonds can be broadly classified into four types; covalent, ionic, van der Waals and metallic. As an example of covalent bonding we shall study the H_2 molecule. Lithium iodide (LiI) molecule will be studied to illustrate ionic bonding. Having described bonding, we shall study some physical properties of molecules. Molecules can rotate like a rigid rotator and vibrate like a spring. These properties are not possessed by atoms.

You may recall that although an electron in a hydrogen atom is held to the nucleus by the attractive coulombic force its motion is restricted to orbits which correspond to quantum mechanical stationary states. The rotational and vibrational motions of molecules are also restricted to configurations which correspond to stationary states. Therefore, the rotational and vibrational energy levels of the molecule are discrete.

A molecule can be excited from its ground state to a stationary state and an excited molecule can make transitions between stationary states to reach its ground state by emitting photons. This process gives rise to rotational and vibrational spectra. When a molecular spectrum is viewed without much resolution, it appears as a series of bands. Each band when resolved finely is seen to consist of closely spaced spectral lines. Physical parameters of molecules such as their interatomic distances and bond strengths can be estimated by analysing the rotational-vibrational spectrum. We shall describe these properties in the following sub- sections.

13.15 Bonding in molecules

13.15.1 Ionic bond

Although all chemical bonds occur because electrons can be placed simultaneously near two nuclei, yet it is often true that electrons

are not shared equally. Sometimes, the electrons though close to both nuclei, tend to be nearer to one nucleus than to the other. When this happens atoms participating in the bonding effectively become oppositely charged and exert on each other a coulombic force of attraction. Such a bonding is called the *ionic bond.*

The alkali halide molecules provide the best examples of the ionic bond. These molecules are diatomic. They are made out of an alkali atom and a halogen atom. When an alkali atom and a halogen atom come close to each other, the weakly bound electron from the alkali atoms shifts to the halogen atom whose electrons now form a stable closed shell configuration. In the process, an ion pair is produced. The positively charged alkali metal ion attracts the negatively charged halide ion and settles down in an equilibirum state at a separation denoted by R, which is the equilibrium bond length. This state corresponds to the lowest energy of the system. The ion pair is prevented from getting closer to each other than the bond length R by the repulsion of electrons in the ionic clouds and the repulsion of the nuclei when they are very close.

Estimates of the dissociation energy of halide molecules can be made using the data on the ionisation potentials, E_i, of alkali atoms and electrons affinity, E_a, of halogen atoms and their internuclear separation R. You may recall that the ionisation potential of an atom is defined to be the minimum energy required to release an electron from it. The ionisation potentials of the alkali metal atoms are as follows:

	E_i(eV)		E_i(eV)
Li	5.40	Rb	4.20
Na	5.18	Cs	3.90
K	4.35		

Similarly, the amount of energy released when a halogen atom acquires an extra electron is called its *electron affinity*. The electron affinities, E_a, of halogen atoms are as follows:

	E_a(eV)
F	−3.45
Cl	−3.62
Br	−3.41
I	−3.06

The experimentally measured value of the dissociation energy of LiI molecule is 3.64 eV. The internuclear separation between Li^+ and I^- ions has been estimated from spectroscopic data to be 2.39 Å. Its dissociation energy can be estimated by following three hypothetical steps.

Let us first assume that 5.40 eV of energy, which is equal to the ionisation potential of Li, has been spent in ionising a Li atom. Let the released electron be captured by an iodine atom liberating 3.06 eV of energy, which is equal to its electron affinity. Hence, the net amount of energy required to form a pair of Li^+ and I^- ions from a pair of Li and I atoms is 2.34 eV (5.40 eV − 3.06 eV). When Li^+ and I^- attain equilibrium at a separation of 2.39 Å their coulombic potential energy is

$$\begin{aligned}&\frac{-e^2}{4\pi\epsilon_0 R}\\ &= \frac{-(1.6\times10^{-19})^2}{4\pi\times 8.85\times10^{-12}\times 2.39\times10^{-10}}\ \mathrm{J}\\ &= -6.02\ \mathrm{eV}.\end{aligned}$$

Therefore, energy equal to 3.68 eV (−6.02 eV + 2.34 eV) will be liberated when Li and I atoms form a lithium iodide molecule. Equivalently, we can say that the total potential energy of a pair of Li and I atoms is 3.68 eV less when they are separated by 2.39 Å from each other, than when they are infinitely separated.

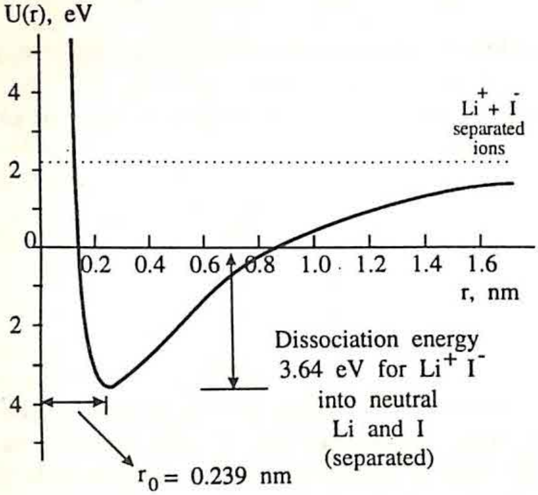


Figure 13.16: Potential energy for Li^+ and I^- ions as a function of separation distance r. The energy at finite separation was chosen to be 2.34eV, corresponding to the energy needed to form ions from neutral atoms. The minimum energy for this curve is at the equilibrium separation $r_0 = 0.239$nm for the ions in the molecule.

You may have noticed that our estimation of dissociation energy of LiI molecule gives a value which is 0.04 eV higher than the experimentally measured value of 3.64 eV. The small difference is a measure of the repulsive interactions mentioned earlier, because Li^+ and I^- ions are not point charges.

Using quantum mechanics, it is possible to estimate the potential energy of the system consisting of Li and I atoms as a function of their separation. The potential energy curve of the system has the appearance of a potential well as shown in Fig. 13.16. It may be noted from this curve that when Li and I atoms are brought closer than equilibrium internuclear separation R, the potential energy sharply increases due to repulsions between electron clouds and between the pair of nuclei. Potential energy curves of this type of shape indicate bonding.

Example 13.7: The equilibrium internuclear distance, R, and dissociation energy, E_d, of the halide molecules of Li and Na are as given below.

	R(Å)	E_d(eV)
LiCl	2.02	4.87
LiBr	2.17	4.39
LiI	2.39	3.64
NaF	1.92	4.95
NaCl	2.36	4.22
NaBr	2.50	3.74

Estimate the repulsive interaction energy contributions to the potential energy of these molecules.

Answer: To answer this problem, we follow the three steps discussed in the text for estimating the dissociation energy for the LiI molecule. The solution has been given in a tabular form (Table 13.2).

We have assumed that the electron is *completely* transferred from the alkali atom to the halogen. In fact the charge transferred is somewhat less than $|e|$; the bond is not *fully* ionic.

13.15.2 Covalent bond

In covalent bonds, as the prefix 'co' implies bonding electrons are shared equally between the two nuclei. Therefore, distinct ions do not take part in this bonding. The best example of covalent bond is the hydrogen molecule H_2. When two hydrogen atoms come sufficiently close to each other there is a simultaneous pull on each electron by both protons with the result that the two electrons are more likely to be found in the space between the two nuclei than elsewhere, for most of the time. The sharing of electrons acts as a glue in cementing hydrogen atoms to form H_2 molecules.

The covalent bond is a result of electri-

Table 13.2: Some data for diatomic molecules with an ionic bond.

Halide molecule	Equilibrium inter-nuclear distance R (Å)	Coulombic potential energy of ion pair $\frac{-e^2}{4\pi\epsilon_0 R} = -\frac{14.38}{R(\text{Å})}$ (eV)	Ionisation potential, E_i, of the participating alkali atom (eV)	Electron affinity E_a, of the participating halogen atoms (eV)	Potential energy of the pair of alkali and halogen atoms at the equilibrium separation R $E_i + E_a - \frac{e^2}{4\pi\epsilon_0 R}$ (eV)	Dissociation energy, E_d (eV)	Contribution of repulsive interactions to the energy of the molecule $\lvert E_i + E_a - \frac{e^2}{4\pi\epsilon_0 R}\rvert - E_d$ (eV)
LiCl	2.02	−7.12	5.40	−3.62	−5.34	4.87	0.47
LiBr	2.17	−6.63	5.40	−3.41	−4.64	4.39	0.25
LiI	2.39	−6.02	5.40	−3.06	−3.68	3.64	0.04
NaF	1.92	−7.49	5.18	−3.45	−5.76	4.95	0.81
NaCl	2.36	−6.09	5.18	−3.62	−4.53	4.22	0.31
NaBr	2.50	−5.75	5.18	−3.41	−3.98	3.74	0.24

cal forces and special quantum mechanical effects. The electron possesses an internal angular momentum which is called its spin. There are two spin states of electron denoted as up (↑) or down (↓). A detailed theoretical calculation based on quantum mechanics shows that when electrons pair with opposite spins such as ↑↓ the energy of the hydrogen atom reduces as atoms come near to each other and leads to bonding. And, when electrons pair with parallel spins ↑↑ or ↓↓ calculation shows that the energy of the hydrogen atoms increases when the atoms come near each other. The result is that this configuration is not energetically favourable for bonding.

13.15.3 van der Waals bond

Atoms of inert gases, which because of their closed-shell electronic structures neither favour ionic nor covalent bondings, exert a small attraction towards each other. It is this attraction which is responsible for the departure of inert gases from the ideal gas law and for their liquefaction at sufficiently low temperatures. van der Waals forces are weak and arise due to induced dipole-induced dipole interactions between atoms. The dipoles arise because electron clouds in atoms are continually in motion. It can happen that sometimes the centre of the electronic charge-cloud does not overlap with the nucleus and a dipole is produced. The dipole interaction energy or the van der Waals attraction energy varies as $1/r^6$, where r is the internuclear separation. Atoms in molecules and in condensed matter have an internuclear separation of about 2-4 Å. Although the energy due to the van der Waals bonding between a pair of atoms is small compared to ionic and covalent bonds, it can become significant in large molecules with many contacts between atoms.

13.15.4 Metallic bond

The existence of metals cannot be explained by any of the types of bonding dealt with so far. Metals have conduction electrons (electrons outside the closed shell) which are free to move around in the solid. They also have positive atomic or ionic residues. It turns out that this collection of an electron liquid and ions is more stable than neutral atoms.

13.16 Molecular energies

As mentioned earlier, molecules have rotational and vibrational spectra. Rotational energies are smaller than the vibrational energies. The rotational spectra lie in the far-infrared and the vibrational spectra lie in the infrared. In Fig. 13.17 we have given the energy level diagram for a molecule undergoing simultaneous rotational and vibrational motions. Such a system is called a vibrating rotator.

In the following sections we have given the theory of the rotational spectrum and the vibrational spectrum of a diatomic molecule using arguments similar to that of Bohr's theory of hydrogen atoms. Although these sections have been starred they can be understood and enjoyed without much difficulty as the mathematics used is restricted to what you know.

*13.17 Rotational spectrum

The rotational motion of a diatomic molecule resembles that of a rigid rotator. A diatomic molecule like LiI has finite size and in the first approximation can be treated as a rigid dumb-bell. It has rotational motion about an axis which passes through the centre of mass of the molecule, perpendicular to the line joining the two ions. You will appreciate that at molecular dimensions the physics of motion can be correctly described

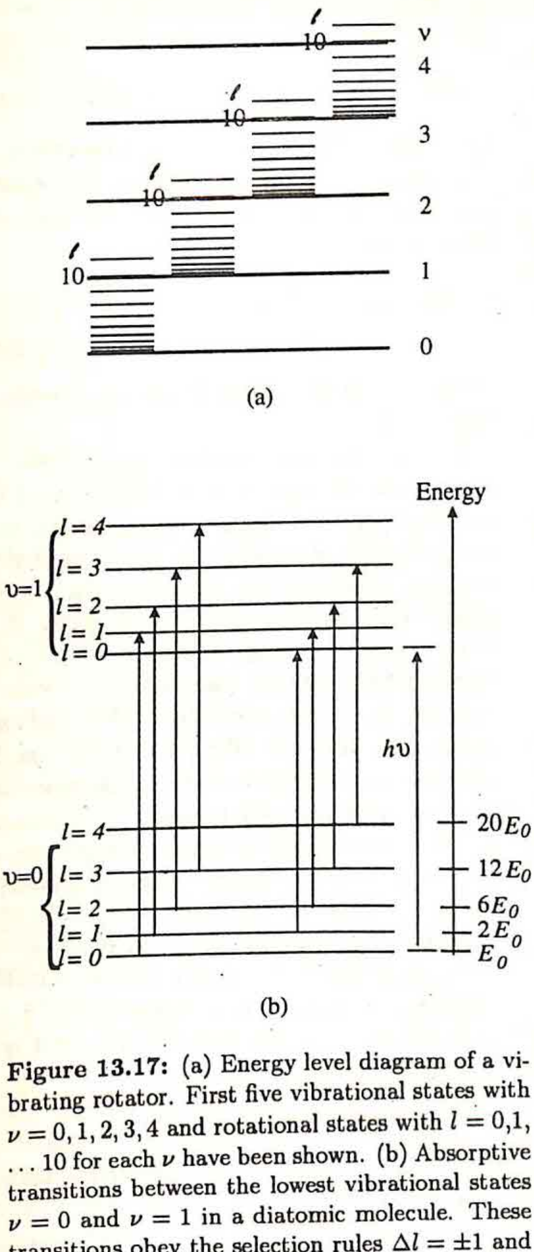


Figure 13.17: (a) Energy level diagram of a vibrating rotator. First five vibrational states with $\nu = 0, 1, 2, 3, 4$ and rotational states with $l = 0,1, \ldots 10$ for each ν have been shown. (b) Absorptive transitions between the lowest vibrational states $\nu = 0$ and $\nu = 1$ in a diatomic molecule. These transitions obey the selection rules $\Delta l = \pm 1$ and fall into two bands. The energies of the $l \to l+1$ band are $h\nu + 2E_0$, $h\nu + 4E_0$, $h\nu + 6E_0$, and so forth; whereas the energies of the $l \to l-1$ band are $h\nu - 2E_0$, $h\nu - 4E_0$, $h\nu - 6E_0$, and so forth. ($E_0 = h^2/8\pi^2 I$).

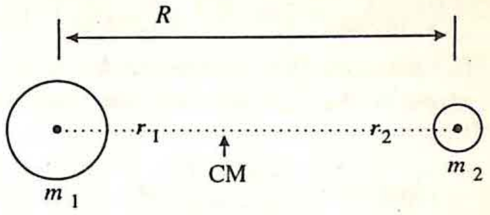


Figure 13.18: Symbolic representation of a diatomic molecule. R is the internuclear separation. r_1 and r_2 are distances of atoms of masses m_1 and m_2 from the centre of mass of the molecule.

only by quantum mechanics. We expect the Planck constant h to make its appearance in the formulae for the energy of the stationary states. We shall, however, treat the molecular rotation classically and put in by hand quantum modifications at appropriate stages.

As stated above, for working out the rotational motion, we approximate diatomic molecule with a massless bar of fixed length with point masses fixed at its two ends. As shown in Fig. 13.18. Let the distances of the ends where ions of masses m_1 and m_2 are held fixed from the centre of mass of the molecule be r_1 and r_2, respectively. The distances r_1 and r_2 are fixed by the relation

$$m_1 r_1 = m_2 r_2 \tag{13.50}$$

and

$$r_1 + r_2 = R \tag{13.51}$$

Hence,

$$r_1 = \frac{m_2 R}{m_1 + m_2}, r_2 = \frac{m_1 R}{m_1 + m_2} \tag{13.52}$$

As the molecule is assumed to be rigidly rotating, its atoms will rotate with the same angular velocity ω. Their speeds of rotation will be $r_1\omega$ and $r_2\omega$, respectively. In classical mechanics, the total rotational energy of the molecule, $(E_{\text{rot}})_{\text{Cl}}$ the sum of the kinetic energy of the two atoms.
That is,

$$(E_{\text{rot}})_{\text{Cl}} = \frac{1}{2}m_1 r_1^2\omega^2 + \frac{1}{2}m_2 r_2^2\omega^2 \quad (13.53)$$

Substituting the expressions for r_1 and r_2 given in Eq. (13.52) you can easily check that

$$(E_{\text{rot}})_{\text{Cl}} = \frac{m_1 m_2 R^2}{2(m_1+m_2)}\omega^2 \quad (13.54)$$

The moment of inertia I of this system about an axis passing through the centre of mass and perpendicular to the line joining m_1 and m_2 can be obtained from its definition, i.e.,

$$\begin{aligned} I &= m_1 r_1^2 + m_2 r_2^2 \\ &= \frac{m_1 m_2}{m_1+m_2}R^2 = \mu R^2, \end{aligned} \quad (13.55)$$

where

$$\mu = \frac{m_1 m_2}{m_1+m_2} \quad (13.56)$$

is called the reduced mass of the two particle system.

Therefore,

$$(E_{\text{rot}})_{\text{Cl}} = \frac{1}{2}I\omega^2 \quad (13.57)$$

The angular momentum, L, of the system

$$L = m_1 r_1^2\omega + m_2 r_2^2\omega = I\omega \quad (13.58)$$

Combining Eqs (13.57) and (13.58) we get

$$(E_{\text{rot}})_{\text{Cl}} = \frac{L^2}{2I} \quad (13.59)$$

You have studied the properties of angular momeutun in Chapter 7 of the class XI physics text book. According to classical mechanics, the angular momentum can take any continuous value. However, quantum mechanics has a big surprise. According to it, the angular momentum of a rigid rotator is quantised. It can take only discrete values. The quantised values of the angular momentum can be labeled by a variable l which takes only integer values 0,1,2,3, ... Also, the Planck constant h, which has the dimension of angular momentum, comes in. The quantum mechanical result for the square of the angular momentum of a rigid rotator is

$$L^2 = l(l+1)\times\frac{h^2}{4\pi^2},\ l = 0,1,2,3,\ldots (13.60)$$

Combining this result with the expression for the energy of the rigid rotator, the energy states of rotation of a diatomic molecule are found to be

$$\begin{aligned} (E_{\text{rot}})_{qm} &\equiv (E_{\text{rot}})_l = \frac{l(l+1)h^2}{8\pi^2 I}, \\ l &= 0,1,2,3,\ldots \end{aligned} \quad (13.61)$$

The rotational energy levels are shown in Fig. 13.19.

If we extend Bohr's hypothesis to molecules, we expect each transition of the molecule from a higher energy state to a lower energy state shall be accompanied by emission of a quantum of electromagnetic energy. This emitted photon will carry away the excitation energy of the molecule. According to quantum mechanics all possible transitions between rotational states of ionic molecules are not allowed. They are restricted by *selection rules*. Ionic molecules possess electric dipole moment. In molecules with electric dipole moment, transitions occur only if the quantum number l changes by one. That is, allowed transitions obey the selection rule $\Delta l = \pm 1$. In other words, a rotating molecule can only make transition from a level with angular momentum $l+1$ to a level with angular momentum l and emit a photon of frequency. ν_l,

$$\begin{aligned} h\nu_l &= (E_{\text{rot}})_{l+1} - (E_{\text{rot}})_l \\ &= \frac{h^2}{8\pi^2 I}[(l+1)(l+2) - l(l+1)] \\ &= \frac{h^2}{4\pi^2 I}(l+1). \end{aligned} \quad (13.62)$$

Also, a molecule can absorb a photon of frequency ν_l and make transition from the state l to the state $l+1$. In molecular spectroscopy it is useful to characterise the spectral lines

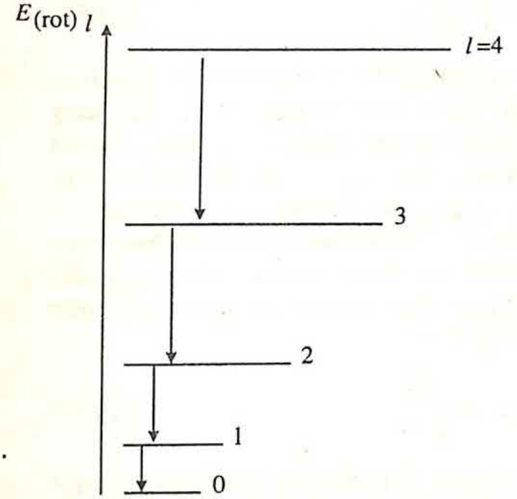


Figure 13.19: Rotational energy levels and some allowed transitions.

with their wave number $\sigma = 1/\lambda$ rather than with their wavelength λ.

The expression for wave number σ_l of rotational spectrum can be obtained by dividing Eq. (13.62) with hc. It gives

$$\sigma_l = \frac{h}{4\pi^2 Ic}(l+1) \qquad (13.63)$$

The wave numbers are equally spaced. This is easily checked. You may note that

$$\sigma_{l+1} - \sigma_l = \frac{h}{4\pi^2 Ic}, l = 0, 1, 2, 3, \ldots (13.64)$$

is independent of l. This is a characteristic property of rotational molecular transitions. The rotational spectral lines lie in the far-infrared and microwave regions.

We give in detail an example to show how from the rotational spectrum, moment of inertia of the molecule and its bond lengths are estimated.

Example 13.8: The rotational spectrum of HCl contains the following wavelengths: 12.03×10^{-5} m, 9.60×10^{-5} m, 8.04×10^{-5} m, 6.89×10^{-5} m, 6.04×10^{-5} m. If the isotopes involved are ${}^1_1H$ and ${}^{35}_{17}Cl$, find the moment of inertia and the distance between hydrogen and chlorine nuclei in a HCl molecule. The mass of ${}^{35}Cl$ is 5.81×10^{-26} kg and that of ${}^1_1H = 1.67 \times 10^{-27}$ kg.

Answer: The wave numbers $\sigma = 1/\lambda$ for the wavelengths of the spectral lines given in the statement of the problem and their differences are easily computed to be

$$\begin{array}{ll}
8.31 \times 10^3\ \mathrm{m}^{-1} & \\
 & = 2.11 \times 10^3\ \mathrm{m}^{-1} \\
10.42 \times 10^3\ \mathrm{m}^{-1} & \\
 & = 2.02 \times 10^3\ \mathrm{m}^{-1} \\
12.44 \times 10^3\ \mathrm{m}^{-1} & \\
 & = 2.07 \times 10^3\ \mathrm{m}^{-1} \\
14.51 \times 10^3\ \mathrm{m}^{-1} & \\
 & = 2.05 \times 10^3\ \mathrm{m}^{-1} \\
16.56 \times 10^3\ \mathrm{m}^{-1} &
\end{array}$$

You can note that within the accuracy of experimental measurements the difference between the successive wave numbers is constant. This observation confirms that the spectral lines belong to rotational spectrum. We can, therefore, use the result given in Eq. (13.64) for estimating the moment of inertia of the molecule. By substituting

$$(\Delta\sigma)_{av} = 2.06 \times 10^3 \mathrm{m}^{-1}$$

in

$$I = \frac{h}{4\pi^2 c(\Delta\sigma)}$$

We get

$$I = \frac{6.62 \times 10^{-34}}{4\pi^2 \times 3 \times 10^8 \times 2.06 \times 10^3}\ \mathrm{kg\ m^2}$$
$$= 27.13 \times 10^{-48}\ \mathrm{kg\ m^2}$$

The reduced mass μ of HCl molecule can be obtained from the definition

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

by substituting $m_1 = 5.81 \times 10^{-26}$ kg and $m_2 = 1.67 \times 10^{-27}$ kg. It gives

$$\mu = \frac{58.1 \times 1.67 \times 10^{-27}}{(58.1 + 1.67)} \text{ kg}$$
$$= 1.62 \times 10^{-27} \text{ kg}$$

It is very close to the mass of the ${}^1_1$H atom. Cl atom is acting like a brick wall! The bond length, R, of the molecule can be obtained from the relation

$$R^2 = \frac{I}{\mu}.$$

It gives

$$R = \left(\frac{27.13 \times 10^{-48}}{1.67 \times 10^{-27}}\right)^{1/2} \text{ m}$$
$$= 1.29 \times 10^{-10} \text{ m}$$
$$= 1.29\text{Å}.$$

Diatomic molecules are not exactly rigid dumbbells. The shape of the potential energy curve of a diatomic molecule near its minimum resembles the potential energy curve of a harmonic oscillator. The molecule in addition to its rotation can execute small oscillations as though the atoms at its ends are connected by a spring instead of being tied to a rigid bar. Like the rotational motion, the oscillatory motion of the molecule has to be treated quantum mechanically. In the following, we give some important features of the vibrational motion of molecules and their spectrum.

*13.18 Vibrational spectrum

The potential energy of a diatomic molecule near its minimum can be approximated by the function

$$E = -E_d + \frac{1}{2}k(r - R)^2 \tag{13.65}$$

where

$$k = \left(\frac{d^2E}{dr^2}\right)_{r=R} \tag{13.66}$$

is a measure of the curvature of the potential energy curve near its minimum. You may recall that in the Chapter 12 of the class XI textbook k has been called the spring constant. It gives a measure of the strength of the bond. The classical angular frequency for simple harmonic oscillations of a system of reduced mass μ tied to a spring of spring constant k is

$$\omega_{\text{osc}} = \sqrt{\frac{k}{\mu}} \tag{13.67}$$

The energy of a system undergoing simple harmonic oscillations is $(1/2)kA^2$, where A is the amplitude of the oscillations. For small oscillations it can take continuous values because A is a real continuous variable.

The application of quantum mechanics to simple harmonic oscillations has another big surprise. The energy of oscillations are quantised and are given by the formula

$$E_{\text{vib}} = (n + 1/2)\frac{h}{2\pi}\omega_{\text{osc}}, \quad n = 0, 1, 2, 3, \ldots \tag{13.68}$$

The energy levels of vibrations can be labelled by a discrete variable , n, which takes values $0, 1, 2, \ldots$. The vibrational energy states are equally spaced. The energy values of the quantised oscillator are

$$\frac{1}{2}\left(\frac{h}{2\pi}\right)\omega_{\text{osc}},\ \frac{3}{2}\left(\frac{h}{2\pi}\right)\omega_{\text{osc}},\ \frac{5}{2}\left(\frac{h}{2\pi}\right)\omega_{\text{osc}},\ \frac{7}{2}\left(\frac{h}{2\pi}\right)\omega_{\text{osc}} \ \ldots \tag{13.69}$$

The spacing between adjacent energy levels is

$$\left(\frac{h}{2\pi}\right)\omega_{\text{osc}}$$

The selection rule for an harmonic oscillator undergoing vibrational changes is

$$\Delta n = \pm 1 \tag{13.70}$$

Applying the selection rule we can obtain the frequency of the vibrational emission:

$$h\nu_{\text{vib}} = (n+1+1/2)\left(\frac{h}{2\pi}\right)\omega_{\text{osc}} - (n+1/2)\left(\frac{h}{2\pi}\right)\omega_{\text{osc}},$$

giving

$$\nu_{\text{vib}} = \frac{1}{2\pi}\omega_{\text{osc}} \tag{13.71}$$

which is also the absorption frequency whatever be the quantum number n of the vibrational state of the molecule.

In reality, the vibration-rotation spectra of a molecule are much more complicated than what we have described. The oscillations are not exactly harmonic because the potential is not exactly $(1/2)k(r-R)^2$. They are modified by anharmonic terms in the potential. We will not treat further the properties of the molecular spectrum because of the limitation of the rules of thumb for taking into account quantum corrections to results obtained from the classical physics.

Example 13.9: HCl has a force constant of 516 Nm^{-1}. Estimate the frquency, ν_{vib} and wavelength λ, of its vibrational line.

Answer: In example 13.7 the reduced mass, μ, of the HCl molecule has been calculated. We have found that

$$\mu = 1.62 \times 10^{-27}\text{kg}$$

ν_{vib} can be calculated from the relation

$$\begin{aligned}\nu_{\text{vib}} &= \frac{1}{2\pi}\omega_{\text{osc}} = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \\ &= \frac{1}{2\pi}\sqrt{\frac{516}{1.62\times 10^{-27}}}\ \text{Hz} \\ &= 8.98\times 10^{13}\text{Hz}\end{aligned}$$

Wave length of the vibrational line

$$\begin{aligned}\lambda &= \frac{c}{\nu_{\text{vib}}} = \frac{3\times 10^8}{8.98\times 10^{13}}\ \text{m} \\ &= 3.3\times 10^{-6}\,\mu\text{m}.\end{aligned}$$

It shows that vibrational spectrum lies in the infrafed.

Example 13.10: From the spectrum of hydrogen gas it is found that the fundamental vibrational transition $\sigma = 0$ to $\sigma = 1$, occurs at $\sigma = 4159.2\text{cm}^{-1}$.
Find the effective spring constant $\left(\frac{d^2E}{dr^2}\right)_{r=R}$ for H_2.

Answer: We know $\omega_{\text{osc}} = \sqrt{\frac{k}{\mu}}$ where

$$\begin{aligned}k &= \left(\frac{d^2E}{dr^2}\right)_{r=R} \\ \mu &= \frac{m_1 m_2}{m_1+m_2} \\ \omega_{\text{osc}} &= 2\pi\nu = 2\pi c\sigma \\ &= 2\pi\times 3\times 10^8 \\ &\quad \times 4.1592\times 10^5\text{s}^{-1} \\ &= 78.399\times 10^{13}\text{s}^{-1}.\end{aligned}$$

Reduced mass μ for the hydrogen molecule will be $\frac{1}{2}m_H$. This is $\frac{1}{2}\times 1.672\times 10^{-27}$ kg. Therefore,

$$\begin{aligned}k &= (\omega_{\text{osc}})^2\mu \\ &= (78.399\times 10^{13})^2\times\frac{1}{2}\times 1.672 \\ &\quad \times 10^{-27}\text{N/m}. \\ &= 514\,\text{N/m}.\end{aligned}$$

Table 13.3: Atomic masses of the elements.

Atomic number	Element	Symbol	Atomic Mass (u)	Atomic number	Element	Symbol	Atomic Mass (u)
1	Hydrogen	H	1.00794	40	Zirconium	Zr	91.22
2	Helium	He	4.00260	41	Niobium	Nb	92.9064
3	Lithium	Li	6.941	42	Molybdenum	Mo	95.94
4	Beryllium	Be	9.01218	43	Technetium	Tc	(98)
5	Boron	B	10.81	44	Ruthenium	Ru	101.07
6	Carbon	C	12.011	45	Rhodium	Rh	102.9055
7	Nitrogen	N	14.0067	46	Palladium	Pd	106.42
8	Oxygen	O	15.9994	47	Silver	Ag	107.8682
9	Fluorine	F	18.998403	48	Cadmium	Cd	112.41
10	Neon	Ne	20.179	49	Indium	In	114.82
11	Sodium	Na	22.98977	50	Tin	Sn	118.69
12.	Magnesium	Mg	24.305	51	Antimony	Sb	121.75
13	Aluminium	Al	26.98154	52	Tellurium	Te	127.60
14	Silicon	Si	28.0855	53	Iodine	I	126.9045
15	Phosphorus	P	30.97376	54	Xenon	Xe	131.29
16	Sulfur	S	32.06	55	Cesium	Cs	132.9054
17	Chlorine	Cl	35.453	56	Barium	Ba	137.33
18	Argon	Ar	39.948	57	Lanthanum	La	138.9055
19	Potassium	K	39.0983	58	Cerium	Ce	140.12
20	Calcium	Ca	40.08	59	Praseodymium	Pr	140.9077
21	Scandium	Sc	44.9559	60	Neodymium	Nd	144.24
22	Titanium	Ti	47.88	61	Promethium	Pm	(145)
23	Vanadium	V	50.9415	62	Samarium	Sm	150.36
24	Chromium	Cr	51.996	63	Europium	Eu	151.96
25	Manganese	Mn	54.9380	64	Gadolinium	Gd	157.25
26	Iron	Fe	55.847	65	Terbium	Tb	158.9254
27	Cobalt	Co	58.9332	66	Dysprosium	Dy	162.50
28	Nickel	Ni	58.69	67	Holmium	Ho	164.9304
29	Copper	Cu	63.546	68	Erbium	Er	167.26
30	Zinc	Zn	65.38	69	Thulium	Tm	168.9342
31	Gallium	Ga	69.72	70	Ytterbium	Yb	173.04
32	Germanium	Ge	72.59	71	Lutetium	Lu	174.967
33	Arsenic	As	74.9216	72	Hafnium	Hf	178.49
34	Selenium	Se	78.96	73	Tantalum	Ta	180.9479
35	Bromine	Br	79.904	74	Tungsten	W	183.85
36	Krypton	Kr	83.80	75	Rhenium	Re	186.207
37	Rubidium	Rb	85.4678	76	Osmium	Os	190.2
38	Strontium	Sr	87.62	77	Iridium	Ir	192.22
39	Yttrium	Y	88.9059	78	Platinum	Pt	195.08

Atomic number	Element	Symbol	Atomic Mass (u)
79	Gold	Au	196.9665
80	Mercury	Hg	200.59
81	Thallium	Tl	204.383
82	Lead	Pb	207.2
83	Bismuth	Bi	208.9804
84	Polonium	Po	(209)
85	Astatine	At	(210)
86	Radon	Rn	(222)
87	Francium	Fr	(223)
88	Radium	Ra	226.0254
89	Actinium	Ac	227.0278
90	Thorium	Th	232.0381
91	Protactinium	Pa	231.0359
92	Uranium	U	238.0289
93	Neptunium	Np	237.0482
94	Plutonium	Pu	(244)
95	Americium	Am	(243)
96	Curium	Cm	(247)
97	Berkelium	Bk	(247)
98	Californium	Cf	(251)
99	Einsteinium	Es	(252)
100	Fermium	Fm	(257)
101	Mendelevium	Md	(258)
102	Nobelium	No	(259)
103	Lawrencium	Lr	(260)

The atomic number is the number of protons in the nucleus. The atomic mass is weighted by isotopic abundance in earth's surface, relative to the mass of the carbon 12 isotope, which has been arbitrarily assigned a mass of 12.00000 atomic mass units (u). Relative isotopic abundances vary considerably for many elements in naturally occuring specimens; commercially available samples may also vary due to undisclosed or inadvertent isotope separation. Numbers in parentheses are mass numbers (the whole number nearest the atomic mass, in u) of the most stable isotope of that element. Adapted fron the Handbook of Chemistry and Physics, 66th Edn. 1985-1986.

Summary

1. *In Thomson's model,* an atom is a spherical cloud of positive charges with electrons embedded in it. Oscillations of electrons about their equilibrium positions give rise to radiations of definite frequencies.

2. *In Rutherford's nuclear model,* most of the mass of the atom and all its positive charge are concentrated in a tiny nucleus, and electrons revolve around it.

   This model emerged from Geiger–Marsden experiments in 1911. A collimated beam of 5.5 MeV α–particles from $^{214}_{83}$Bi was allowed to fall on a 2.1×10^{-7} m thin gold foil. The scattered α–particles produced scintillations on a ZnS screen, which were counted at different angles (θ) from the direction of the beam. Though most of the α–particles suffered negligible deviation, some (about one in 10^4) suffered a large change in direction ($\theta > 90°$). The last observation was a crucial clue to the nuclear model.

3. Rutherford's calculation used the inverse-square law of repulsive force between an α–particle ($Z = 2$) and a gold nucleus ($Z = 79$). Multiple–scattering was ignored. The scattering angle θ of the α–particle is related to the impact parameter b by the relation

$$b = \frac{Ze^2 \cot(\frac{\theta}{2})}{4\pi\epsilon_0 \left(\frac{1}{2}mv^2\right)}$$

   (impact parameter is the perpendicular distance of the initial velocity vector of the α–particle from the centre of the nucleus). The observed number of scattered α–particles at different angles agreed with Rutherford's calculation based on the nuclear model of atom.

4. Classically, Rutherford's model of the atom is unstable. An orbiting electron accelerates continuously and must lose energy as electromagnetic radiation. The orbit should shrink spiraly into the nucleus within 10^{-8} s, and give out a continous spectrum of radiation. But we know hydrogen atom is stable and has a characteristic line spectrum.

5. *Bohr's model of the hydrogen atom* introduced radically new postulates as follows and laid the foundation of quantum mechanics:

   (a) In a hydrogen atom, there exists certain special orbits (called stationary orbits) in which electrons do not radiate.

   (b) The stationary orbits are those for which the angular momentum of the electron (L) is an integral multiple of $h/2\pi$. (Bohr's quantization condition)

   (c) When a H–atom makes a transition from a state of energy E_i to another state of emergy E_f, the difference in energy is carried away by a photon of frequency ν_{if} such than

$$h\nu_{if} = E_i - E_f$$

Postulate (b) is equivalent to saying that in a stationary state, the circumference of a circular orbit contains integral number of de Broglie wavelengths.

$$2\pi r = n\lambda = n\frac{h}{mv} \text{ i.e., } L = mvr = n\frac{h}{2\pi}$$

6. For a circular orbit

$$\frac{mv^2}{r} = \frac{e^2}{4\pi\epsilon_0 r^2}$$

The quantization condition on angular momentum gives

$$v = \frac{nh}{2\pi mr}; n = 1, 2, 3, 4, \ldots\ldots$$

These equations give

$$r = \frac{n^2}{m}\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\epsilon_0}{e^2}$$

$$v = \alpha c/n$$

where

$$\alpha = \frac{e^2}{4\pi\epsilon_0 c(h/2\pi)} = \frac{1}{137}$$

is known as the fine-structure constant. Total energy

$$E = \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r}$$

which using earlier formulae fpr v amd r give:

$$E_n = -\frac{1}{2}\frac{mc^2\alpha^2}{n^2} = -\frac{13.6}{n^2}\text{eV}$$

This formula holds good even for elliptic orbits.

The Rydberg formula for the H–atom spectrum is given by

$$\nu_{if} = \frac{E_{n_i} - E_{n_f}}{h} = \frac{1}{2}\frac{mc^2\alpha^2}{h}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

Lyman series (in uv region) corresponds to

$$n_f = 1; \quad n_i = 2, 3, 4, \ldots\ldots$$

Balmer series (in visible region) corresponds to

$n_f = 2; \quad n_i = 3, 4, 5, \ldots\ldots$

7. Beryllium nuclei bombarded with α–particles emit neutral radiation. Using energy–momentum conservation, Chadwick concluded in 1932 that this radiation consists not of photons but of neutral particles we call neutrons. The mass of neutron is obtained from scattering experiments with targets of different masses. $m_n = 1.00866$ u, where 1 u is defined to be mass of ^{12}C atom divided by 12.

8. In the notation ^A_ZX for a nuclear species, X stands for the chemical symbol of the species, Z is atomic number (= no. of protons) and A is mass number (= no. of protons + no. of neutrons). Isotopes of an element have the same Z (and hence identical chemical properties) but have different A. For example, naturally occurring chlorine is a mixture of two isotopes $^{35}_{17}$Cl and $^{37}_{17}$Cl with relative abundances of 75.4% and 24.6%, respectively. The mass of natural chlorine atom is the weighted average of the masses of these two isotopes. The isotopes of the hydrogen are $^1_1$H, $^2_1$H (deutrium), and $^3_1$H(tritium).

9. Neutrons and protons are bound in a nucleus by the short range strong nuclear force. The nuclear force does not distinguish between neutron and proton; which are distinguished by electromagnetic force.

   Mass defect is the difference in mass of a nucleus from the sum of the masses of its constituents. From Einstein's energy–mass relation $E = mc^2$, *Binding energy* of a *nucleus* ^A_ZX is c^2 times its mass defect.

   $$B.E = [Zm_H + (A - Z)m_n - m]c^2$$

   where m_{H} is the mass of a hydrogen atom and m is the mass of the *atom* ^A_ZX. In the range $A = 30$ to $A = 170$, BE per nucleon is nearly constant, about 8 MeV/nucleon.

10. *Nuclear sizes* have been accurately measured by high energy electron scattering experiments. The empirical relation between the radius R of a nucleus and its mass number A is: $R = R_0A^{1/3}$, where $R_0 = 1.1 \times 10^{-15}$ m

    which shows that density of nuclear matter is independent of A. It is of the order of 10^{17} kg m^{-3}.

11. An unstable nuclear state may reach state of greater stability by the emission of an α–particle ($^4_2$He nucleus), or a β–particle (electron) or γ–ray.

    $\alpha - decay: \quad ^A_ZX \longrightarrow {}^{A-4}_{Z-2}Y + {}^4_2\mathrm{He}$

    e.g., $^{222}_{86}\mathrm{Rn} \longrightarrow {}^{218}_{84}\mathrm{Po} + {}^4_2\mathrm{He}$

    The energy Q released in the process is

    $$Q = \left(m_x - m_y - m_{\mathrm{He}}\right)c^2$$

From energy–momentum conservation it is seen that kinetic energy of α–particles:

K $E_{He} \simeq (A-4)\,Q/A$

$\beta - decay$: $^{A}_{Z}X \longrightarrow {}^{A}_{Z+1}Y + e^{-}$ (observed)

As for α-decay, we expect electrons to come out with fixed energy. But experimentally, β–particles have a continuous energy spectrum from 0 to a maximum value. To account for this, Pauli postulated that another particle (now called antineutrino) is also emitted in β–decay which shares the energy of the decay:

$$^{A}_{Z}X \longrightarrow {}^{A}_{Z+1}Y + e^{-} + \overline{\nu}$$

e.g., $^{60}_{27}Co \longrightarrow {}^{60}_{28}Ni + e^{-} + \overline{\nu}$

$\gamma - decay$: When an excited nucleus decays to a lower state, e.m. radiation of very short wavelength (fraction of a Å) is emitted. For example, the excited state of $^{60}_{28}Ni$ (obtained in the β–decay of $^{60}_{27}Co$) reaches the ground state emitting γ–rays (photons), of energies 1.17 MeV and 1.33 MeV.

12. *Radioactive decay law:*

$$\frac{dN}{dt} = -\lambda N(t) \quad \text{i.e.,} \quad N(t) = N(0)e^{-\lambda t}$$

where λ is the probablity per unit time for a nucleus to decay, and $N(t)$ is the number of radioactive nuclei present at time t. Half–life $T_{1/2}$ is the time interval in which one half the number of nuclei decay.

$$T_{1/2} = \frac{\log 2}{\lambda} = \frac{0.693}{\lambda}$$

The decay rate $R(t) = \lambda N(t)$ is measured in units of curie.

1 Ci [curie] $= 3.70 \times 10^{10}$ disintegrations /s.

13. *Nuclear fission:* splitting of heavy nucleus ($A > 230$) when excited into intermediate mass nuclei. For examble

$$^{235}_{92}U + n \longrightarrow {}^{144}_{56}Ba + {}^{89}_{36}Kr + 3n + Q$$

where Q is about 200 MeV.

Nuclear fusion: fusion of light nuclei into a heavy nucleus with energy liberation; e.g., fusion of hydrogen nuclei into the helium nucleus is the source of energy in all stars. These reactions require extremely high temperatures ($10^6 - 10^7$K) so that the nuclei can overcome their electrostatic repulsion.

14. *Nuclear Reactor:* An average of 2.5 neutrons are released per fission of uranium. So a chain reaction is possible. In an atomic bomb, this chain reaction is uncontrolled; a nuclear reactor employs controlled chain reaction rate to produce energy. Slow neutrons have much greater probability of inducing fission of $^{235}_{92}U$. But neutrons

produced in a fission reaction are fast (average energy $\sim$ 2 MeV.) The fast neutrons are slowed down by elastic scattering with light nuclei called Moderators. (e.g. water, heavy water and graphite). The reaction rate is controlled by neutron absorbing material. (e.g., cadmium rods)

Exercises

13.1 What is the distance of closest approach when a 5.0 MeV proton approaches a gold nucleus?

13.2 A 12.5 MeV α-particle approaching a gold nucleus is deflected back by 180^o. How close does it approach the nucleus?

13.3 Two energy levels in an atom are separated by 2.3 eV. What is the frequency of radiation emitted when the atom transits from the upper to lower level?

13.4 State the basic postulates of Bohr's model of hydrogen atom. Derive a formula for the stationary energy levels of a hydrogen atom.

13.5 Give the Rydberg formula for hydrogen spectrum, and state the transitions corresponding to Lyman series and Balmer series.

13.6 The ground state energy of hydrogen atom is -13.6 eV. What are the kinetic energy and potential energy of the electron in this state?

13.7 Using Bohr's formula for the stationary energy levels of a hydrogen atom, obtain the longest wavelength in the Lyman series of hydrogen spectrum.

13.8 Show that, in Bohr's model, radii of electronic orbits increase as n^2, where n is the quantum number of the orbit.

13.9 The radius of the innermost electronic orbit of a hydrogen atom is 5.3×10^{-11}m. What are the radii of the $n = 2$ and $n = 3$ orbits?

13.10 Obtain the binding energy of a nitrogen nucleus from the following data: $m_H = 1.00783$ u, $m_n = 1.00867$ u, $m({}^{14}_{7}N) = 14.00307$ u. Give your answer in units of MeV. [Remember $1u = 931.5 MeV/c^2$].

13.11 From the relation $R = R_0 A^{\frac{1}{3}}$, where R_0 is a constant and A is the mass number of a nucleus, show that the nuclear matter density is nearly constant. (i.e. independent of A).

13.12 Write the nuclear equation for:

(i) the α-decay of ${}^{226}_{88}Ra$

(ii) the β^- decay of ${}^{32}_{15}P$

(iii) the β^+ decay of ${}^{11}_{6}C$

13.13 A radioactive isotope has a half-life of T years. After how much time is its activity reduced to 6.25% of its original activity?

13.14 The half-life of ${}^{90}Sr$ is 28 years. How much is the disintegration rate of 15 mg of this isotope?

13.15 Obtain the amount of ${}^{60}Co$ necessary to provide a radioactive source of 8.0 Ci strength. The half-life of ${}^{60}Co$ is 5.3 years.

13.16 The equilibrium internuclear distance of a NaCl molecule is 2.36 Å.

Estimate the Coulomb potential energy of the ion pair.

13.17 Explain briefly the function of a 'moderator' in a nuclear reactor.

Additional Exercises

13.18 Choose the correct alternative from the clues given at the end of each statement:

(a) The *size of the atom* in Thomson's model is ...... the atomic size in Rutherford's model. (much greater than, no different from, much less than)

(b) In the ground state of ...... electrons are in stable equilibrium, while in ...... electrons always experience a net force. (Thomson's model, Rutherford's model).

(c) A *classical* atom based on ...... is doomed to collapse. (Thomson's model, Rutherford's model)

(d) An atom has a nearly continuous mass distribution in ...... but has a highly non-uniform mass distribution in ...... (Thomson's model, Rutherford's model.)

(e) The positively charged part of the atom possesses most of the mass of the atom in ...... (Rutherford's model, both the models.)

13.19 Answer the following questions which will help you understand the difference between Thomson's model and Rutherford's model more precisely.

(a) Is the average angle of deflection of α-particles by a thin gold foil predicted by Thomson's model much less, about the same, or much greater than that predicted by Rutherford's model?

(b) Is the probability of backward scattering (i.e. scattering of α-particles at angles greater than 90°) predicted by Thomson's model much less, about the same, or much greater than that predicted by Rutherford's model?

(c) Keeping other factors fixed, it is found experimentally that for a small thickness t, the number of α-particles scattered at moderate angles is proportional to t. What clue does this linear dependence on t provide?

(d) In which model is it completely wrong to ignore multiple scattering for the calculation of average angle of scattering of α-particles by a thin foil?

[Note: Detailed proofs of answers given at the end are beyond the scope of this book. You need only know them as important features which help you discriminate between the two models].

13.20 A projectile of mass m, charge Z, initial speed v and impact parameter b is scattered by a heavy nucleus

of charge Z. Use angular momentum and energy conservation to obtain a formula connecting the minimum distance s of the projectile from the nucleus to these parameters. Take $b = 0$, and obtain a formula for the distance of closest approach r_0, [Ignore size of the nucleus and its recoil motion].

13.21 For the scattering of α-particles (of energy in the range of a few MeV) by a gold foil about 1μm thick, calculations based on Thomson's model (beyond our scope here) predict the average angle of deflection $<\theta>$ to be about $1°$ in fair agreement with experiments. The probability P of scattering at angles greater than θ is given in this model to be $P(\geq \theta) = e^{-\theta^2/<\theta>^2}$ (You can take this formula as given without proof). Estimate the probability of backward scattering in Thomson's model. [When you compare your theoretical estimate with the observed result that $P(\geq 90°) \simeq 10^{-4}$, you will appreciate the real failure of Thomson's model].

13.22 For scattering by an 'inverse-square' field (such as that produced by a charged nucleus in Rutherford's model) the relation between impact parameter b and the scattering angle θ is given by:

$$b = \frac{Ze^2 \cot(\theta/2)}{4\pi\epsilon_0(mv^2/2)}$$

(a) What is the scattering angle for $b = 0$?

(b) For a given impact parameter b, does the angle of deflection increase or decrease with increase in energy?

(c) What is the impact parameter at which the scattering angle is $90°$ for $Z = 79$ and initial energy equal to 10 MeV?

(d) Why is it that the mass of the nucleus does not enter the formula above but its charge does?

(e) For a given energy of the projectile, does the scattering angle increase or decrease with decrease in impact parameter?

13.23 The gravitational attraction between electron and proton in a hydrogen atom is weaker than the coulomb attraction by a factor of about 10^{-40}. An alternative way of looking at this fact is to estimate the radius of the first Bohr orbit of a hydrogen atom if the electron and proton were bound only by gravitational attraction. You will find the answer interesting!

13.24 Using Bohr's formula for energy quantization, determine (i) the excitation energy of the $n = 3$ level of the He^+ atom (ii) the ionization potential of the ground state of Li^{++} atom.

13.25 Obtain a formula for the frequency of radiation emitted when a hydrogen atom de-excites from level n to level $(n-1)$. For large n, show that this frequency equals the classical frequency of revolution of the electron in the orbit.

[This result is an example of an important principle due to Niels Bohr (Bohr's Correspondence Principle). For large quantum numbers, quantum mechanical results generally reduce to classically expected results].

13.26 Which state of the triply ionized beryllium (Be^{+++}) has the same orbital radius as that of the ground state of hydrogen? Compare the energies of the two states.

13.27 Which level of the doubly ionized lithium (Li^{++}) has the same energy as the ground state energy of the hydrogen atom? Compare the orbital radii of the two levels.

13.28 The total energy of an electron in the first excited state of the hydrogen atom is about -3.4eV.

(a) What is the kinetic energy of the electron in this state?

(b) What is the potential energy of the electron in this state?

(c) Which of the answers above would change if the choice of the zero of potential energy is changed?

13.29 Determine the speed of the electron in the $n = 3$ orbit of He^{+}. Is the non-relativistic approximation valid?

13.30 If Bohr's quantization postulate (angular momentum $= nh/2\pi$) is a basic law of nature, it should be equally valid for the case of planetary motion also. Why then do we never speak of quantization of orbits of planets around the sun?

13.31 Classically, an electron can be in any orbit around the nucleus of an atom. Then what determines the typical atomic size? Why is an atom not, say a thousand times bigger or smaller than its typical size? This question had greatly puzzled Bohr before he arrived at his famous model of the atom that you have learnt in the text. To simulate what he might well have done before his discovery, let us play as follows with the basic constants of nature and see if we can get a quantity with the dimension of length that is roughly equal to the known size of an atom($\sim 10^{-10}$m).

(a) Construct a quantity with the dimension of length from the fundamental constants, e, m_e and c. Determine its numerical value.

(b) You will find that the length obtained in (a) is many orders of magnitude smaller than the atomic size. Further it involves c. But energies of atoms are mostly in non-relativistic domain where c is not expected to play any role. This is what may have suggested Bohr to discard c and look for 'something else' to get the right atomic size. Now the Planck's constant h had already made its appearance elsewhere. Bohr's great ingenuity lay in recognising that h, m_e and e will yield the right atomic size. Construct

a quantity with the dimension of length from e, m_e and h; and confirm that its numerical value has indeed the correct order of magnitude.

13.32 A helium atom consists of two electrons orbiting round a nucleus of charge $Z = 2$. But the electrons do not 'see' the full charge $Z = 2$ of the nucleus. Each electron sees the nucleus slightly 'screened' by the other electron, so that the effective charge Z_{eff} seen by each electron is less than 2. The ionization potential for a helium atom in its ground state is measured experimentally to be 24.46 V. Estimate the effective charge of the nucleus as seen by each electron in the helium ground state.

13.33 Obtain the first Bohr's radius and the ground state energy of a 'muonic hydrogen atom' (i.e. an atom in which a negatively charged muon (μ^-) of mass about 207 m_e orbits around a proton).

13.34 A pi-mesic hydrogen atom is a bound state of a negatively charged pion (denoted by π^-, $m_{\pi^-} = 273$ m_e) and a proton. Estimate the number of revolutions a π^- makes (averagely) in the ground state of the atom before it decays (mean life of a $\pi^- \simeq 10^{-8}$s).

13.35 A positronium atom is a bound system of an electron (e^-) and its antiparticle, the positron (e^+) revolving about their centre of mass. In which part of the em spectrum does the system radiate when it deexcites from its first excited state to the ground state?

13.36 Give the order of magnitude of nuclear mass density and average atomic mass density. Compare these densities with the typical mass density of solids, liquids and gases (at ordinary temperatures and pressures).

13.37 Obtain approximately the ratio of the nuclear radii of the gold isotope $^{197}_{79}$Au and the silver isotope $^{107}_{47}$Ag. What is the approximate ratio of their nuclear mass densities?

13.38 The three stable isotopes of neon: ^{20}Ne, ^{21}Ne, ^{22}Ne have respective abundances of 90.51%, 0.27% and 9.22%. The atomic masses of the three isotopes are 19.99 u, 20.99 u and 21.99 u, respectively. Obtain the average atomic mass of neon.

13.39 (a) From 13.38 you have learnt that atomic mass of an element is the weighted average of the atomic masses of different isotopes of that element. This explains why atomic masses of many elements show large departures from integer values. However, even if we consider masses of individual isotopes, they are not strictly integer multiples of the mass of a hydrogen ($^1_1$H) atom. How do you account for this fact?

(b) The isotope $^{16}_{8}$O has 8 protons, 8 neutrons and 8 electrons, while $^{8}_{4}$B has 4 protons, 4 neutrons and 4 electrons. Yet the ratio of their atomic masses is not exactly 2. Why?

13.40 The binding energy of a nucleus ^A_ZX is given by the formula

$$B.E. = [Zm_H + (A-Z)m_n - m(^A_ZX)]C^2$$

where m(^A_ZX) is the atomic mass of X. Derive this equation; state clearly the approximation involved and say why it is a very safe approximation.

13.41 Obtain the binding energy of the nuclei $^{56}_{26}$Fe and $^{209}_{83}$Bi in units of MeV from the following data:

$$m_{\rm H} = 1.007825\text{u},$$
$$m_n = 1.008665\text{u},$$
$$m(^{56}_{26}\text{Fe}) = 55.934939\text{u},$$
$$m(^{209}_{83}\text{Bi}) = 208.980388\text{u},$$

which nucleus has greater binding energy per nucleon?

13.42 The neutron separation energy is defined to be the energy required to remove a neutron from a nucleus. Obtain the neutron separation energies of the nuclei $^{41}_{20}$Ca and $^{27}_{13}$Al from the following data:

$$m_n = 1.008665\text{u},$$
$$m(^{40}_{20}\text{Ca}) = 39.962591\text{u},$$
$$m(^{41}_{20}\text{Ca}) = 40.962278\text{u},$$
$$m(^{26}_{13}\text{Al}) = 25.986895\text{u},$$
$$m(^{27}_{13}\text{Al}) = 26.981541\text{u},$$

13.43 The nucleus $^{238}_{92}$U is unstable against α-decay with a half-life of about 4.5×10^9 years. Write down the equation of the decay and estimate the kinetic energy of the emitted α-particles from the following data:

$$m(^{238}_{92}\text{U}) = 238.05081\text{u},$$
$$m(^4_2\text{He}) = 4.00260\text{u},$$
$$m(^{234}_{90}\text{Th}) = 234.04363$$

13.44 (a) If the α-decay of ^{238}U is energetically allowed (i.e. the decay products have a total mass less than the mass of ^{238}U), what prevents ^{238}U from decaying all at once? Why is its half-life so large?

(b) The α-particle faces a Coulomb barrier (see answer to (a) above). A neutron being uncharged faces no such barrier. Why does the nucleus $^{238}_{92}$U not decay spontaneously by emitting a neutron?

$$m(^{238}_{92}\text{U}) = 237.04874\text{u},$$
$$m_n = 1.00867\text{u}.$$

13.45 The nucleus ^{23}Ne decays by β^- emission. Write down the β-decay equation and determine the maximum kinetic energy of the electrons emitted from the following data:

$$m(^{23}_{10}\text{Ne}) = 22.994466\text{u}$$
$$m(^{23}_{11}\text{Na}) = 22.089770\text{u}.$$

13.46 (a) The *observed* decay products of a free neutron are a proton and an electron. The emitted electrons are found to have a continuous distribution of kinetic energy with a maximum of $(m_n - m_p - m_e)c^2$. Explain clearly why the presence of a continuous distribution of energy is a pointer to the existence of other unobserved product(s) in the decay.

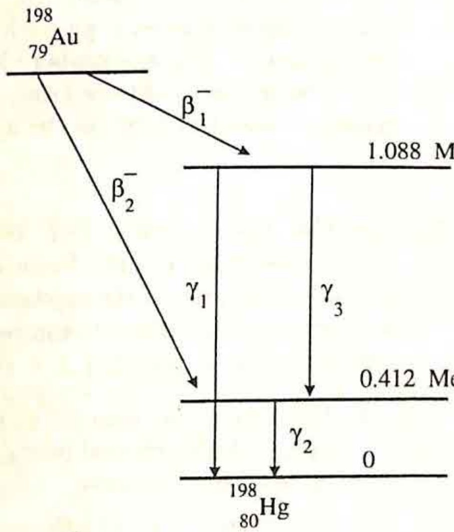
(b) If a neutron is unstable with a half-life of about 917s, why don't all the neutrons of a nucleus decay eventually into protons? How can a nucleus of Z protons and $(A-Z)$

neutrons ever remain stable if the neutrons themselves are unstable?

13.47 For the β^+ (positron) emission from a nucleus, there is another competing process known as electron capture (electron from an inner orbit (say the K-shell) is captured by the nucleus and a neutrino is emitted) $e + {}^A_Z X \longrightarrow {}^A_{Z-1} Y + \nu$). Show that if β^+ emission is energetically allowed, electron capture is necessarily allowed but not vice versa.

13.48 Obtain the maximum kinetic energy of β-particles, and the radiation frequencies corresponding to γ-decays in the following decay scheme. You are given that

$$m(^{198}\text{Au}) = 197.968233\text{u}$$
$$m(^{198}\text{Hg}) = 197.966760\text{u}$$



13.49 The normal activity of living carbon-containing matter is found to be about 15 decays per minute for every gram of carbon. This activity arises from the small proportion of radioactive ^{14}C present with the ordinary carbon isotope ^{12}C. When the organism is dead, its interaction with the atmosphere (which maintains the above equilibrium activity) ceases and its activity begins to drop. From the known half life (= 5730 years) of ^{14}C, and the measured activity, the age of the specimen can be approximately estimated. This is the principle of ^{14}C dating used in archaeology. Suppose a specimen from Mohenjodaro gives an activity of 9 decays per minute per gram of carbon. Estimate the approximate age of the Indus-Valley civilization. (Note, the example is artificial, meant only to give you practice with the principle of the ^{14}C dating method).

13.50 The Q-value of a nuclear reaction A + B → C + D is defined by Q $= [m_A + m_B - m_C - m_D]c^2$ where the masses here refer to *nuclear* rest masses. Determine from the given data whether the following reactions are exothermic or endothermic.

(i) $p + {}^3_1\text{H} \rightarrow {}^2_1\text{H} + {}^2_1\text{H}$

(ii) ${}^{12}_6\text{C} + {}^{12}_6\text{C} \rightarrow {}^{20}_{10}\text{Ne} + {}^4_2\text{He}$

Atomic masses are given to be

$$m_H = 1.007825\text{u},$$
$$m({}^2_1\text{H}) = 2.014102\text{u},$$
$$m({}^3_1\text{H}) = 3.016049\text{u},$$
$$m({}^{12}_6\text{C}) = 12.000000\text{u},$$
$$m({}^{20}_{10}\text{Ne}) = 19.992439\text{u},$$
$$m({}^4_2\text{He}) = 4.002603\text{u},$$

13.51 Answer the following questions:

(a) Are the equations of nuclear reactions (such as those given in 13.50) 'balanced' in the sense a chemical equation (e.g. $2H_2+O_2 \rightarrow 2H_2O$) is? If not, in what sense are they balanced on both sides?

(b) If both the number of protons and the number of neutrons are conserved in each nuclear reaction, in what way is mass converted into energy (or vice versa) in a nuclear reaction?

(c) A general impression exists that mass-energy interconversion takes place only in nuclear reactions and never in chemical reactions. This is, strictly speaking, incorrect. Explain.

13.52 Consider one of the fission reactions of ^{235}U by thermal neutrons:

$$^{235}_{92}\mathrm{U} + \mathrm{n} \rightarrow ^{94}_{38}\mathrm{Sr} + ^{140}_{54}\mathrm{Xe} + 2\mathrm{n}$$

The fission fragments are, however, not stable. They undergo successive β^- decays until $^{94}_{38}$Sr becomes $^{94}_{40}$Zr and $^{140}_{54}$Xe becomes $^{140}_{58}$Ce. Estimate the total energy released in the process. Is all that energy available as kinetic energy of the fission products (Zr and Ce)? You are given that $m(^{235}\mathrm{U}) = 235.0439$ u, $m_n = 1.00866$ u, $m(^{94}\mathrm{Z}) = 93.9065$ u, $m(^{140}\mathrm{Ce}) = 139.9055$ u.

13.53 Suppose India has a target of producing by 2000 AD. 100,000 MW of electric power, ten percent of which is to be obtained from nuclear power plants. Suppose we are given that, on an average, the efficiency of utilization (i.e. conversion to electric energy) of thermal energy produced in a reactor is 25%. How much amount of fissionable uranium would our country need per year at the turn of this century? Take the heat energy per fission of ^{235}U to be about 200 MeV. Avogadro's number $= 6.023 \times 10^{23}$ mol^{-1}.

13.54 (a) In a nuclear reactor explain clearly the role of

(i) moderator (why is heavy water used as a moderator)

(ii) control rods (why are they made of cadmium)

(iii) delayed neutrons

(b) Safety of nuclear reactors is an important issue that has attracted much attention recently. Guess some of the safety problems that a nuclear engineer must cope with in reactor design. [Note, a proper view of this issue can be obtained only by reading a standard text on the subject.]

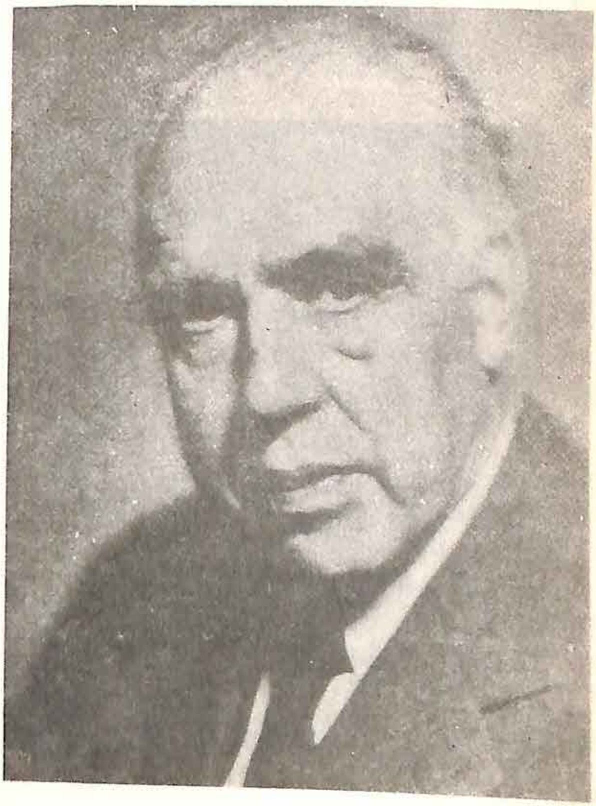
13.55 Consider the so called D-T reaction (deuterium - tritium fusion reaction) which may be the reaction of a future thermo-nuclear fusion reactor. $^2_1\mathrm{H} + ^3_1\mathrm{H} \rightarrow ^4_2\mathrm{He} + \mathrm{n}$

(a) Determine the amount of energy in MeV released in the reaction from the data:

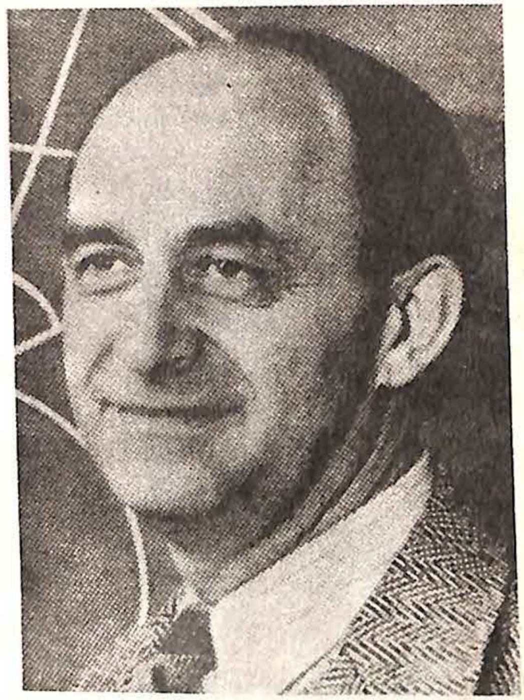
$$m(^2_1\mathrm{H}) = 2.014102\mathrm{u},$$
$$m(^3_1\mathrm{H}) = 3.016049\mathrm{u},$$
$$m(^4_2\mathrm{He}) = 4.002603\mathrm{u},$$
$$m_n = 1.008665\mathrm{u},$$

(b) Consider the radius of both deuterium and tritium to be approximately 1.5×10^{-15}m. What is the kinetic energy needed to overcome Coulomb repulsion? To what temperature must the gases be heated to initiate the reaction?

Rutherford, Ernest (1871-1937) British physicist who did pioneering work on radioactive radiation. He discovered alpha-rays and beta-rays. Along with Frederick Soddy, he created the modern theory of radioactivity. He studied thorium emanation which led to the discovery of the noble gas thoron. By scattering alpha-rays from thin metal foils, he discovered the atomic nucleus and proposed the planetary model of atom. He also estimated the approximate size of the nucleus.



Bohr, Niels Henrik David (1885-1962) Danish physicist who explained the spectrum of hydrogen atom based on quantum ideas. He gave a theory of nuclear fission based on the liquid-drop model of nucleus. Bohr contributed to the clarification of conceptual problems in quantum mechanics, in particular by proposing the complementarity principle.



Fermi, Enrico (1901-1954) Italian physicist who was both a theorist and an experimentalist. He studied nuclear transmutations and showed that radioactive isotopes of most elements could be produced by neutron bombardment. This work resulted in the discovery of nuclear fission and the production of elements lying beyond what was until then the periodic table. He was responsible for the development of the atomic pile and the first controlled nuclear chain reaction. Fermi discovered the quantum statistical laws governing the particles appropriately called the fermions.

Curie, Maire née Skladowaka (1867-1934) Born in Poland, recognised both as a physicist and as a chemist. The discovery of radioactivity by Henri Becquerel in 1896 inspired Marie and her husband Pierre Curie in their researches and analyses which led to the isolation of radium and polonium elements. She was the first person to be awarded two Nobel Prizes - for Physics in 1903 and for Chemistry in 1911.

CHAPTER 14

Solids and Semiconductor Devices

14.1 Introduction

In this chapter, we describe very briefly some properties of solids, and discuss at length some applications of semiconducting solids to electronic devices. In Section 14.2, we begin with a description of solids and their spatial structure. Many properties of a solid depend on the electrons in the outermost shell, e.g. on how they are arranged in energy and in space and how they interact with each other. This *electronic structure* of solids is briefly alluded to in Section 14.3. An example is their classification into metals, insulators and semiconductors. Semiconductors and their properties are also introduced here. The next section (Section 14.4) describes in some detail the pn junction *diode*, which consists of two kinds of (doped) semiconductor joined together. This device has a much larger resistance for electric current flow in one direction than in the opposite, and is therefore used for converting alternating current into direct current. This process is known as rectification. A number of rectifier circuits are discussed and other applications of pn junction diodes pointed out.

The emphasis in this section is on the use of the diode as a circuit element with given current voltage characteristics and not on the detailed physical understanding of how it works since the latter requires more solid state physics and quantum mechanics than

we have done. Section 14.5 describes the transistor, a semiconductor device that has revolutionized the modern world. The transistor and combinations of it as an amplifier, and as an oscillator are some simple illustrative examples discussed. Typical circuits using transistors are analyzed. The final Section 14.6 is on digital electronics, which uses the effect of combinations of diodes etc. on electrical pulses to realize logical functions. This is the operational basis of digital computers, for example.

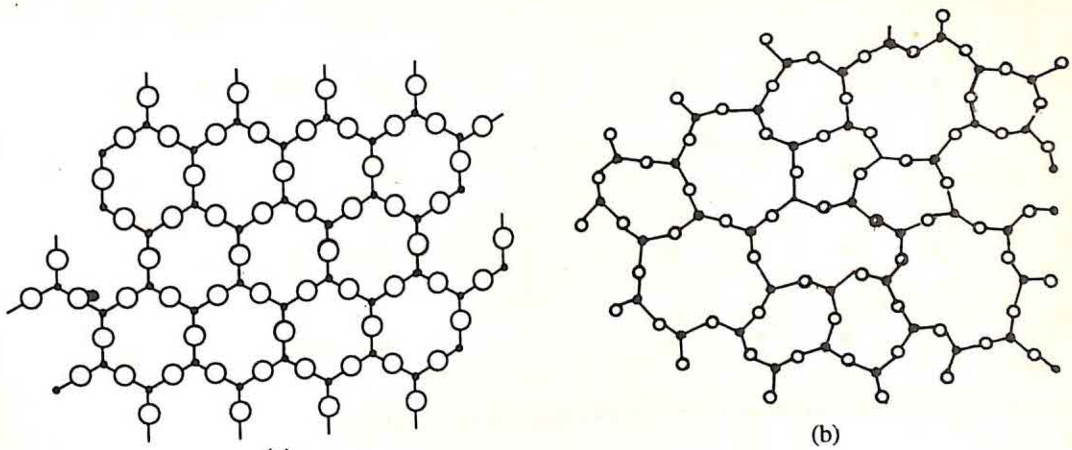
14.2 Solids and their structure

14.2.1 Emergence of new properties in condensed matter

When a large number of atoms and molecules are close together, they condense to form a liquid, or a solid. Since the electrons of different atoms are near each other, in the condensed phase they affect each other. This gives rise to new kinds of collective electronic behaviour which often does not even make sense for the constituent or individual atoms. Also, there are many ways in which collections of atoms can arrange or organize themselves. These ways depend on the nature of atoms and interactions between them, and external conditions such as temperature, pressure etc.. There is an infinite variety of systems that can be put together. For all these reasons, the physics of condensed matter has proved very rich, exhibiting a bewildering variety of unexpected phenomena. One consequence is the large variety of applications. Another is that new physical concepts and theories are needed to make sense of observations. These ideas in turn have applications outside physics, such as in chemistry, biology and materials science. We shall not describe in any depth this large and rapidly expanding area of physics (now called condensed matter physics), but mention a few examples to illustrate what has been said above. We then outline the elements of the structure of solids, before going on to a description of semiconductors, which constitute one class of solids.

An example of how new properties emerge is provided by superconductivity. For example, when niobium atoms are put together to form a solid, the solid is a hard metal. On cooling to below 9K, it completely losses all electrial resistance, becoming a *superconductor.* There is no change in its structure as one cools through 9K, but some subtle reorganization does take place in the motion of electrons in niobium which causes such a fundamental change in its electrical behaviour. Even more surprising is the occurrence of superconductivity above liquid air temperature in $YBa_2Cu_3O_7$, an oxide. Generally, oxides eg. Y_2O_3, BaO, CuO are good insulators. Why is this compound a metal and what makes it a superconductor at such relatively high temperature? Some oxides are ferroelectric like $BaTiO_3$, showing a spontaneous electric dipole moment below a certain temperature and some others like quartz are piezoelectric, i.e. develop an electric dipole moment under stress. The latter property is used in piezoelectric oscillators, in quartz watches for example. Some metals, like Fe, Co and Ni, become spontaneously magnetic (ferromagnetic) below a certain temperature. What is the physical origin of these properties? What are the new phenomena exhibited by such systems? Why do some systems show these properties and not others? These are some natural questions that arise.

Another class of systems, much in use by nature, is soft; large scale changes in the arrangement of parts can take place in them at not much cost in energy. Examples are polymers, colloids and membranes. These systems show unusual behaviour with say small



(a) (b)

Figure 14.1: Arrangement of two-atom groups in (a) a crystal and (b) an amorphous material.

stresses or with temperature.

We cannot go into any of these unusual effects or questions here, but focus on the most obvious feature that distinguishes solids from other states of matter, namely the fact that solids have a definite *shape*.

14.2.2 Structure of solids

It was guessed at long back that the definite shape of a solid could be due to the parts constituting it being arranged in a regular and fixed way, well before atoms were clearly identified as building blocks of matter. Many solids have well define facets or faces (e.g., crystals of sugar or common salt). This suggested to many scientists that solids consits of identical units arranged in a periodic manner. In such a case, there will be some definite planes or faces between regularly arranged units. For example, when a crystal grows in a constant environment, the shape often remains unchanged during growth, as if identical elementary building blocks were being added continuously to it. We now know that these units or building blocks are atoms, or groups of atoms. Crystals are a three dimensional periodic array of atoms. A particular solid (element or compound) is identified by a specific arrangment of atoms, called crystal structure. We mention some general types of crystal structure below, and describe the ideas invovled. But before that, we make the point that not all solids are crystalline.

Consider a hypothetical system with two species of molecules A and B, in the ratio two to three. The atoms are such that it is energetically favoured for an atom A to have three B atoms as neighbours, and for the atom B to have two A atoms as neighbours, in both cases with fixed interatomic distances. How does a collection of atoms A and B arrange itself in space, under these conditions? There is an infinite number of arrangements, two of which in a plane, are shown in Fig. 14.1. The first is a regular, periodic, arrangement of A_2B_3 with the same pattern of atom positions repeated throughout the plane (Fig. 14.1a). The second has the same local neighbourhood, namely three B atoms next to each A atom and two A atoms next to each B atom at the same distances as in the crystal, but no periodicity or regularity. Both the first and second arrangements have *short range order*, namely at short distances they are ordered (e.g. A and B are with respect to each other as near-

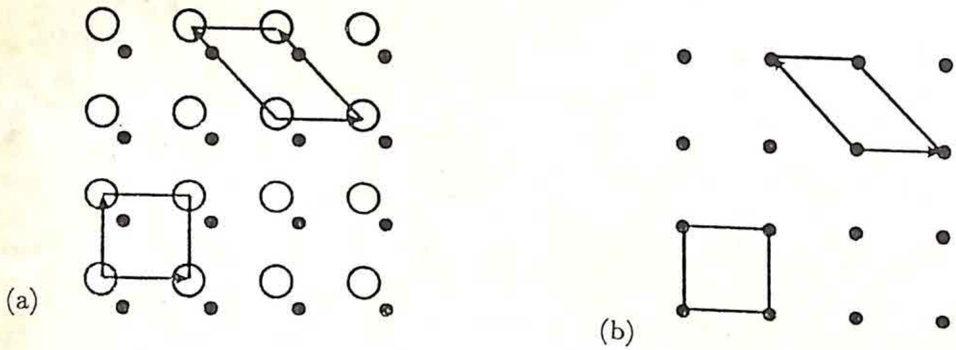


Figure 14.2: (a) Arrangement of two-atom groups in a crystal and (b) its crystal lattice and unit cell.

est neighboures). Only the first has *long range order*, namely has a given local arrangment repeated *identically* to arbitrarily large distances. The first is a crystal, and the second is an amorphous or glassy solid. There is an infinite number of possible glassy structures. In reality, the same substance can exist in both forms, depending conditions! For example, if molten silica (SiO_2) is cooled slowly, it becomes crystalline quartz at low temperature. If it is cooled rapidly, the atoms do not have enough time to move into the right places and it becomes glassy (vitreous silica). The cooling rates required to produce glassy structures can vary by several orders of magnitude; simple systems cannot generally be cooled fast enough to make them glassy. We shall not consider glassy systems here, though they are very common and have unusual properties many of which are basically not understood. We proceed with the description of crystal structure.

14.2.3 Crystal structure

In order to describe the structure of a crystal, it is clearly not necessary to give the positions of all atoms in it.

We can take the hypothetical example of a structure in two dimensions. The two atom unit represented by an open circle and a filled circle repeats periodically, Fig. 14.2(a). We can select a unit consisting of a group of atoms, called the *unit cell*, which repeats itself to form the entire crystal. The unit cell can be chosen in a different ways; two possible unit cells are shown in Fig. 14.2(b). The structure of a crystal is determined completely by the unit cell, its constituents and its manner of repetition.

The replacement of a crystal by a *lattice* simplifies the study of geometrical arrangement of atoms in a crystal. The crystal lattice in the above case is obtained as follows. First, we associate a point with each two atom group and then replace all the groups by points at corresponding positions. This gives an array of points which is known as *crystal lattice*, Fig. 14.2(b). Each point of the lattice is called a lattice point or *lattice site*. This two dimensional picture can be extended to three dimensions. For example Fig. 14.3(a) shows how the entire three dimensional crystal can be obtained by identically and periodically repeating the same unit cell many times in three different directions in space, Fig. 14.3(b).

The lengths a, b, c of a cell are called the *lattice constants* of a crystal. Thus the unit cell is a parallelopiped constructed by the

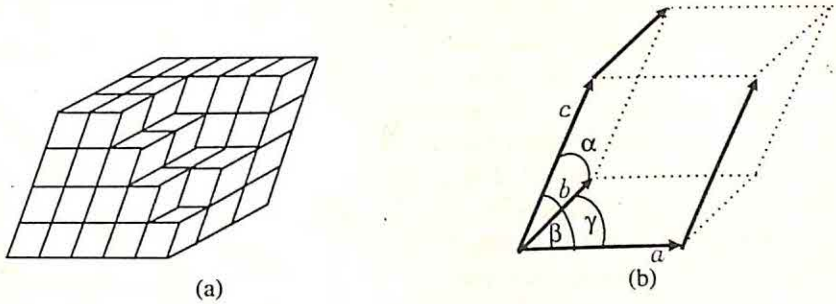


Figure 14.3: Three dimensional representation of arrangement of unit cells in a crystal. (a) Building of the crystal using unit cells as building blocks and (b) Representation of unit cell by lattice constants a, b, c and angles between them α, β and γ.

three vectors **a, b, c**. The unit cell is also characterised by the angles between these vectors. These angles are designated by α, β and γ (see Fig. 14.3). The crystal structure is fully defined by lattice constants a, b, c and angles α, β, γ. With respect to the shape of the unit cell all crystals can be grouped into seven crystallographic systems (Table 14.1).

We illustrate these systems by some examples. The simplest and fairly common structures are cubic and hexogonal. Fig. 14.4a shows the unit cell of a face centred cubic crystal. The atoms are located at the corners of the cube, and at the centres of the faces. This for example is the structure of copper, that is, copper atoms are located as shown. This is also the structure of NaCl. This structure is *close packed*, i.e. if the atoms, assumed identical and spherical, are packed as densely as possible (with the spheres touching each other), the crystalline or periodic structures that result are face centred cubic (fcc) and hexagonal close packed (hcp). We notice that each atom has *twelve* nearest neighbours, another indication of close packing. The body centred cubic is another common cubic structure where the atoms are at the cubes edges, and at the cube centre. Elements like sodium, chromium, and iorn have the bcc structure, as also the salt Cs Cl. Fig. 14.4b,c show the

Table 14.1: The fourteen lattice types in three-dimensions.

System	Restrictions on conventional unit-cell axes and angles
Triclinic	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$
Monoclinic	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$
Orthorhombic	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
Cubic	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
Trigonal	$a = b = c$ $\alpha = \beta = \gamma < 120^\circ, \neq 90^\circ$
Hexagonal	$a = b \neq c$ $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$

bcc structure. In Fig. 14.4b, a bcc unit cell is shown, and CsCl is shown in Fig. 14.4c. The hexagonal close packed or hcp structure is shown in Fig. 14.5. Elements like Mg, Be, and Zn have hcp structure. This structure is best imagined as follows. Consider equal size spheres on a plane, forming a close packed structure. Clearly this has hcp symmetry; each sphere is surounded by six other spheres which it touches. The next close packed layer of spheres is placed so that the spheres of this layer lie in the triangular depression due to packed spheres in the first plane (Fig. 14.5). A very common structure is that of Si and Ge. This can be thought of a two interpenetrating fcc structures (Fig. 14.6). We notice that this structure has four nearest neighbours, and is thus quite open.

14.2.4 Electronic structure of solids

You have already learnt about atomic structure in Chapter 13. We know from the electronic structure of the hydrogen atom that the energies of the electrons in an atom cannot have arbitrary values but only some definite values given by quantum mechanical laws. The electrons have well defined energy levels for a free atom. However, if the atom belongs to a crystal, where it is surrounded by neighbouring atoms, then the energy levels are modified. While this modification is not appreciable in the case of the energy levels of electrons in the inner shells, it is considerable in the case of the energy levels of the electrtons in the outermost shells since these are the electrons which are shared by more than one atom in the crystal.

It turns out that the energy levels of an electron in a solid consist of *bands* of allowed states. There are regions of energy, called gaps, where no states are possible. In each allowed band, the energy levels are very closely spaced, the spacing between succes-

(a)

(b)

(c)

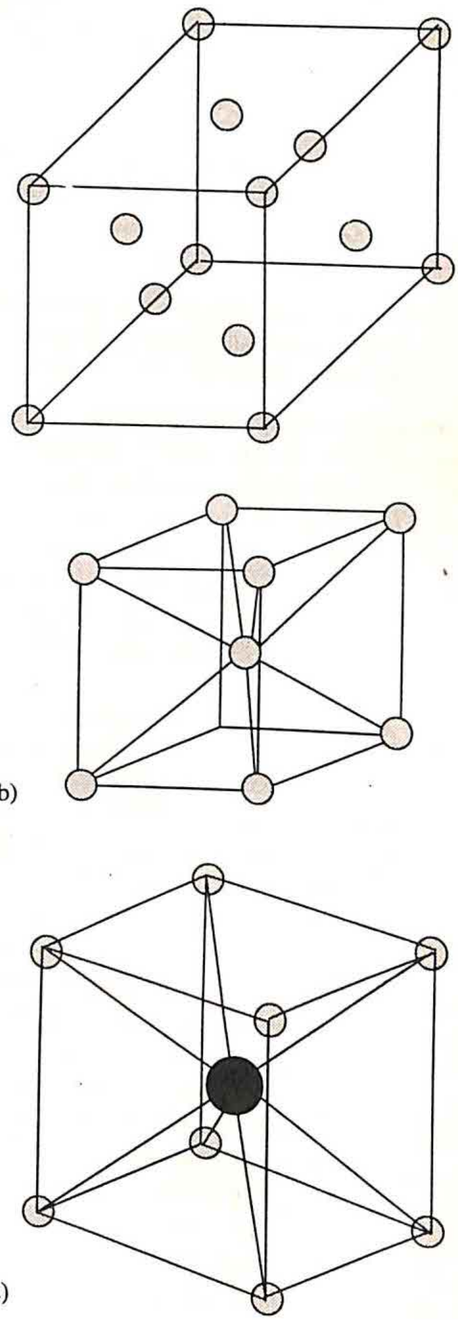
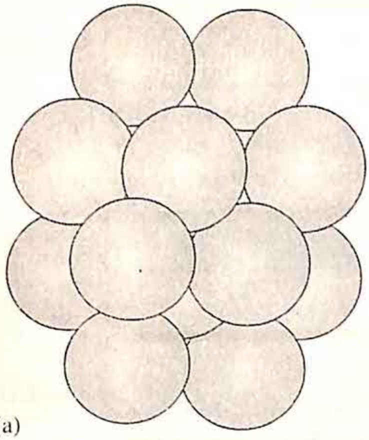


Figure 14.4: Some common crystal structures. (a) unit cell of a face centered cubic crystal; (b) bcc unit cell and (c) Arrangement of Cs^+ and Cl^- in crystalline CsCl.



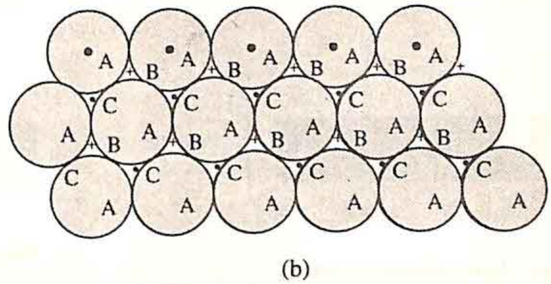


Figure 14.5: (a) Close packing of spheres and (b) the hexagonal close packed (hcp) structure.

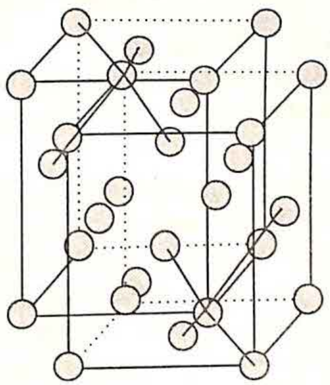


Figure 14.6: Diamond structure; Si and Ge have this crystal structure.

sive levels being of order $(\hbar^2/mV^{2/3})$ where $\hbar$ is Planck's constant $h/2\pi$, m is the electron mass, and V the volume of the solid. This band structure of electronic energies in a solid is shown in Fig. 14.7. The width of a band at typical densities depends on the nature of the wave function of the outer shell electrons, and is typically of the order of a few electron volts. The number of electronic states in a band turns out to be $2N$ where N is the number of atoms in the crystal. This fact is important, because electrons obey the Pauli exclusion principle, namely there can be only one electron in one state (quantum state). This principle is to be combined with the fact that electrons occupy states which minimize the total energy. For noninteracting electrons the total energy is the sum of the energies of the individual electrons. Thus electrons occupy one by one, the lowest energy states till the last electron in the system occupies the state with the highest energy. Thus, depending on the number of electrons, and on the arrangement of the bands, a band may be fully occupied (filled band) or partially occupied. Depending upon the energy band structure, a solid can be placed in one of the three categories - metal, insulator or semiconductor. Metals are good conductors of electricity, insulators are very poor conductors of electricity, and the conductivity of semiconductors lies between that of metals and insulators on a logarithmic scale.

Metals

The energy band structure of a metal is shown schematically in Fig. (14.7(a)). The last occupied band of energy levels is only partially filled; the available electrons occupy, one by one (Pauli exclusion principle) the lowest levels. This leaves part of the

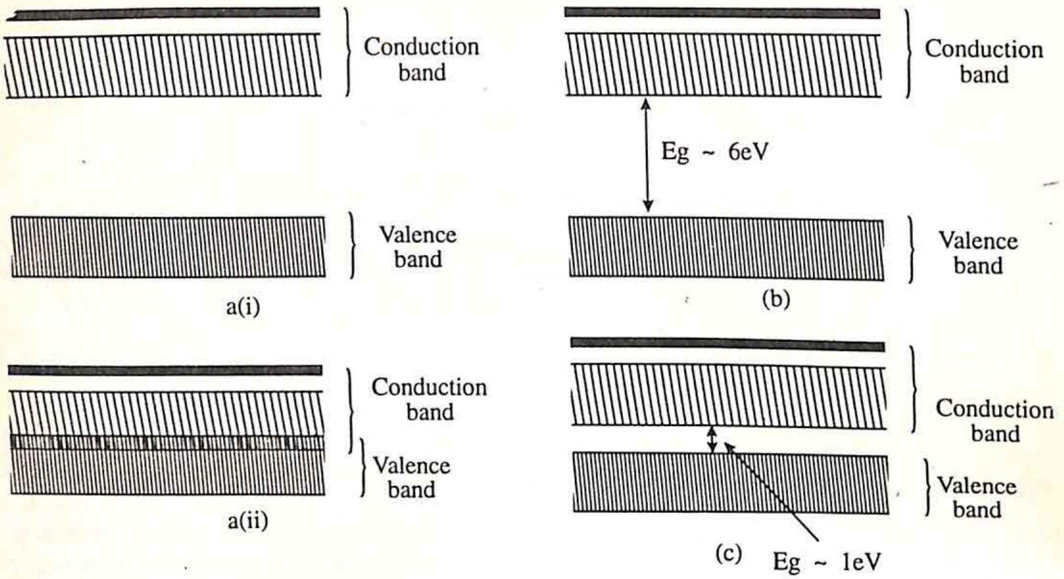


Figure 14.7: Energy band diagram for a (a) metal, (b) insulator, and (c) semiconductor: Note that one can have a metal either when the conduction band is partially filled or when the conduction and valence bands overlap in energy.

band (called conduction band) unoccupied. There are two broad arrangements of bands, both of which lead to a metal. In one there is a gap (energy gap) between the completely filled valence band and the partially filled conduction band (Fig. 14.7(a(i))). The partial or incomplete filling of the latter is the reason why the solid is a metal. The second possibility is shown in Fig. 14.7(a(ii)). The valence band is full and the conduction band empty, but the two overlap in energy. This is called a band overlap metal. Why does such a solid conduct electricity? Consider a weak electric field E applied to the solid. The electrons are accelerated, i.e. gain energy. For this to be possible, this state with additional energy should be unoccupied. Now, as discussed in Chapter 3, if l is the mean free path, the typical energy gain is eEl. You can work out this energy - it is of order 10^{-8} eV for typical laboratory fields! So, one needs unoccupied states infinitesmally close in energy to the occupied states. In such a case, the process of conduction described in Chapter 3, namely acceleration, collision and retun to equilibrium, leading to a steady electron drift, can occur. If the highest occupied electronic state is separated from the lowest unoccupied state by an energy gap, the system is an insulator.

Insulators

The energy band structure of an insulator is shown in Fig. 14.7(b). It shows that the conduction band is separated from the valence band by a wide energy gap (which is, e.g. 6 eV for a diamond crystal). When an electric field is applied, conduction is impossible as discussed above. Therefore, insulators are very poor conductors of electricity. However, at any nonzero temperature, some electrons (a fraction $p \propto \exp(-E_g/k_BT)$ according to Boltzmann's law) can be excited to the conduction band, which is now not fully empty. These electrons can therefore conduct elec-

tricity, with a conductivity proportional to p.

The distinction between insulators and semiconductor is quantitative, not qualitative. (Both of them have an energy gap separating occupied and unoccupied states). It depends on the size of the gap. If the gap is 1 eV or so, we say the solid is a semiconductor. If the gap is of order 5 eV, we say it is an insulator. The carrier density and hence the conductivity depends exponentially on the ratio of the energy E_g to the temperature (room temperature). This ratio varies from about 30 for a semiconductor to about 150 for an insulator. Correspondingly, the room temperature resistivities are $10^{-3}\Omega$ m and $10^{10}\Omega$ m (A good metal has a room temperature resistivity of order 10^{-8} to $10^{-7}\Omega$ m).

Semiconductors

The energy band strcture of a semiconductor is shown in Fig. 14.7(c). It is similar to that of an insulator but with a comparatively small energy gap. At absolute zero of temperature, the conduction band of semiconductors is totally empty, and all the energy states in the valence band are filled. The absence of electrons in the conduction band at absolute zero does not allow current to flow under the influence of an electric field. Therefore, they are insulators at low temperatures.

However, at room tempertures, some valence electrons acquire thermal energy greater than the energy gap E_g and move to the conduction band where they are free to move under the influence of even a small electric field. As mentioned earlier, this fraction is $p \propto \exp(-E_g/k_BT)$. Since E_g is small, this fraction is sizeable for semiconductors. Thus, a crystal orginally an insulator at low tempertures becomes slightly con-

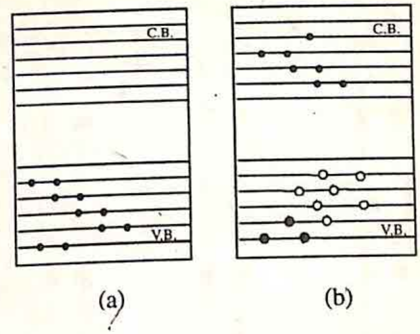


Figure 14.8: Energy band diagram of an intrinsic semiconductor showing some charge carriers at (a) absolute zero, and (b) at room temperature. At room temperature some electrons (•) are in conduction band leaving an equal number of holes (o) in valence band. Note that each horizontal line represents an energy level. In each energy level there can be at most two electrons.

ducting at room temperature. We therefore, call such crystals as *semiconductors*. Unlike metals the resistance of semiconductors decreases with increasing temperature.

14.3 Semiconductors

In this section, we discuss some properties of semiconductors necessary for describing devices made out of them. We begin with a brief introduction to the notion of holes, which are positively charged carriers in semiconductors and metals. We then describe the process called doping, which can greatly affect the electrical properties of semiconductors. Finally, we try to understand qualitatively the simplest semiconductor device, the pn junction rectifier.

14.3.1 Holes

As discussed earlier, some of the electrons in a semiconductor move to the conduction band at high temperatures. These are the *intrinsic carriers*. In the valence band a vacancy is created at the place where the

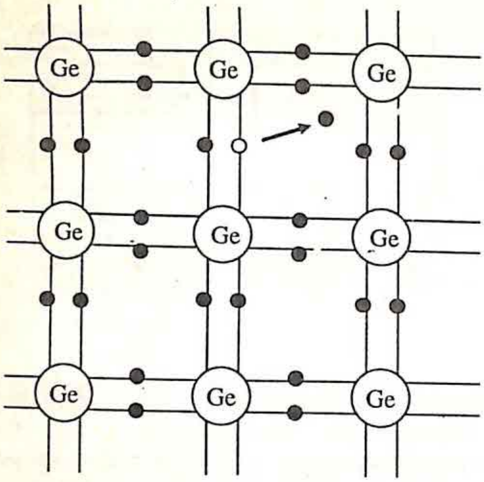


Figure 14.9: Breaking of a covalent bond in a Ge crystal.

electron was present before moving to the conduction band (see Fig. 14.8). This vacancy is called a *hole*.

The way a hole is produced can also be understood by referring to Fig. 14.9. On receiving an additional energy, one of the electrons contributing to a covalent bond breaks and is free to move in the crystal lattice. While coming out of the covalent bond, it leaves behind a hole which is shown as an open circle.

An electron from a neighbouring atom can break away and can come to the place of the missing electron (or hole) completing the covalent bond and creating a hole at another place. We see in our two dimensional example that an electron from any of the eight neighbouring atoms can come to complete the bond and the hole can move to any of these atoms. We can say that the hole moves randomly in a crystal lattice.

★ In a real situation, there is a large number of electrons and holes, and the completion of a bond may not be necessarily due an electron from a bond of a neighbouring atom filling up the vacancy, but it can also be due to a free electron (the conduction band electron) combining with the vacancy i.e. electron-hole recombination. The breaking of bonds or generation of electron-hole pairs, and completion of bonds due to recombination is taking place all the time. At equilibrium, the rate of generation becomes equal to the rate of recombination, giving a fixed number of free electrons and holes. ★

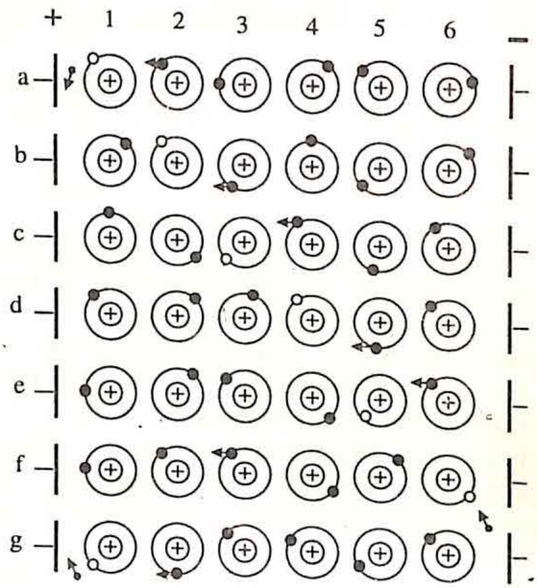


Figure 14.10: Movement of a hole to atoms (1,2,3,$\cdots$ 6) under an electric field at different instants of time (a,b,c, $\cdot\cdot$,g)

We now try to see how the holes move under an applied electric field. This can be understood by an oversimplified picture given in Fig. 14.10 which shows 6 atoms of a semiconductor at 7 instants of time.

In Fig. 14.10, a semiconductor system has been considered in which there are no free electrons and all the electrons are bound with covalent bonds. Suppose by thermal ex-

Table 14.2: Properties of Si and Ge at 300 K.

	Si	Ge
Energy gap E_g(eV)	1.1	0.7
Electron mobility $\mu_n(\mathrm{m^2V^{-1}s^{-1}})$	0.135	0.39
Hole mobility μ_p $(\mathrm{m^2V^{-1}s^{-1}})$	0.048	0.19
Intrinsic carrier concentration n_i $(\mathrm{m^{-3}})$	1.5×10^{16}	2.4×10^{19}
Intrinsic conductivity σ $(\mathrm{Sm^{-1}})$	4.4×10^{-4}	2.18
Intrinsic resistivity $(\Omega\mathrm{m})$	2300	0.46
Density $(\mathrm{gm^{-3}})$	2.3×10^6	5.32×10^6
Concentration of atom $(\mathrm{m^{-3}})$	5×10^{28}	4.41×10^{28}

citation a bond is broken at atom-1, which generates an electron-hole pair. The atom-1 is close to the positive terminal of the battery which collects this electron, while the hole remains at atom-1. Under the action of the electric field an electron from atom-2 goes to atom-1. We say that the hole has moved to atom-2. Similarly, an electron from atom-3 moves to atom-2 and the hole moves to atom-3. In this way, the hole moves to atom-6 which takes an electron from the negative terminal of the battery and the covalent bond is completed. Thus, the electron initially lost to the positive terminal of the battery has been finally compensated by the negative terminal of the battery which constitutes a current.

We notice the following things. Holes have a positive charge, because they move in a direction opposite to that of electrons, for the same electric field. A hole is a convenient way of describing charge motion, though the motion can be described entirely in terms of electrons. For example, Fig. 14.10, can be discussed either in terms of the movement of one hole, or in terms of the movement of five bound electrons on six atoms. Obviously the former is more convenient. Finally, the current in the outside circuit is completed by electron flow as is seen from the above example.

When both electrons and holes are present, and are, for away from each other, we can assume that they carry current independently. The total current is

$$I = I_e + I_h \text{ (or } I_n + I_p). \tag{14.1}$$

To remove an electron from the filled valence band (thus creating a hole) requires more energy, the farther down the electron state is from the top of the valence band. Thus we say that the hole farther from the top of the valence band has higher energy, just as an electron in the conduction band, farther from its bottom has higher energy. Finally, the hole has the same mass as the (removed) electron.

14.3.2 Intrinsic semiconductors

Pure semiconductors are called *intrinsic semiconductors*. In intrinsic semiconductors,

$$n_e^{(0)} = n_h^{(0)} = n_i \tag{14.2}$$

where $n_e^{(0)}$ is the electron density in conduction band, $n_h^{(0)}$ is the hole density in valence band, and n_i is the intrinsic carrier concentration. For Si and Ge, the values of n_i are given in Table 14.2. It is difficult to make an intrinsic semiconductor because of the difficulty in preparing extremely pure materials. Usually, if the impurity concentrations is less than n_i then the semiconductor is called intrinsic.

Table 14.2 includes a quantity called mobility, and given the symbols μ_n and μ_p (for

electrons and holes, respectively). Mobility is the drift velocity v_d acquired by a charge carrier in an electric field E, divided by the electric field. We know from Eq. (3.10) that

$$v_d = (eE\tau/m)$$

so that

$$\mu = (v_d/E) = (e\tau/m). \tag{14.3}$$

From the definition, Eq. (14.3) the units of μ are seen to be $(\mathrm{m/s})/(\mathrm{V/m}) = \mathrm{m}^2\ \mathrm{V}^{-1}.\ \mathrm{s}^{-1}$.

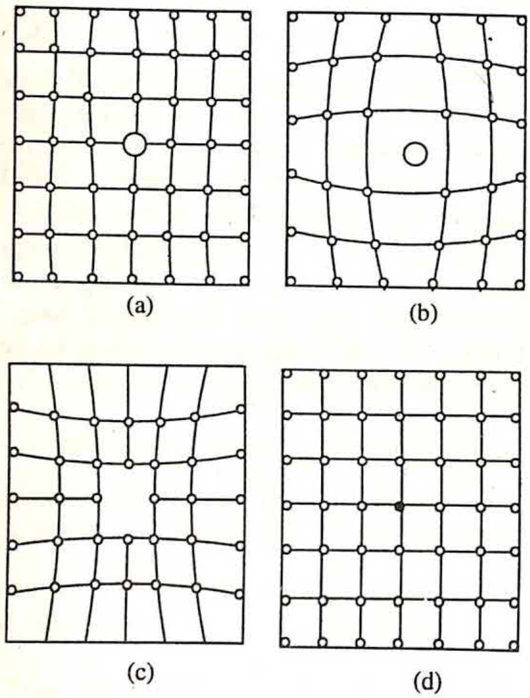


Figure 14.11: Crystal defects due to (a) an impurity atom marked (0) at lattice site, (b) impurity atom marked (0) between lattice sites, (c) missing atom, and, (d) atom (•) added by doping.

14.3.3 Doping a semiconductor

We have underlined the fact that a crystal is a periodic, ordered arrangement of atoms and molecules. However, there are several kinds of *defects in crystals* and it is impossible to have a crystal without any defects. The properties of crystals are influenced by these defects. It is difficult to get 100% chemically pure material. So the crystals grown will have some undesirable impurity atoms at the lattice sites as shown in Fig. 14.11(a,b,c). The atom of the crystal itself may be missing from the lattice site as shown in Fig. 14.11(c). This is called a *point defect.* The undesirable impurity atoms can be removed by chemically purifying the material.

In semiconductor device applications, we require very high purity (99.9999% or more) of materials to start with, so that their electrical properties can be modified in a controlled manner by adding suitable impurities. This deliberate addition of a desirable impurity is called *doping* and the impurity atoms added are called *dopants.* The dopant should, preferably substitute the semiconductor atom (Fig. 14.11(d)) i.e. go into a lattice site occupied originally by the crystal lattice atom. It should not distort the crystal lattice, i.e. its size should be almost the same as that of the crystal atom. The concentration of dopant atoms should not be large (not more than about 1% of the crystal atoms). It may be emphasised that the effect of doping can be seen only in extremely pure samples. With the doping of a semiconductor, the conductivity is greatly enhanced as discussed below.

In practice, doping is achieved in many ways. One is to add the impurity atoms in the melt of the semiconductor. Another is to heat the crystalline semiconductor in an atmosphere containing dopant atoms or molecules containing dopant atoms, so that the latter diffuse into the semiconductor. A third is to implant dopant atoms by bombarding the semiconductor with their ions.

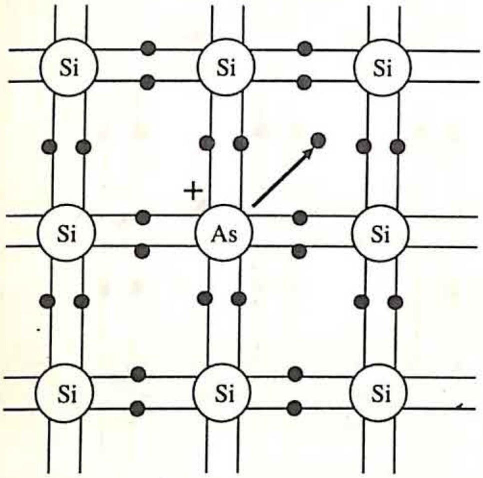


Figure 14.12: Sharing of electrons in covalent bonds, with a V-group atom present (n-type semiconductor).

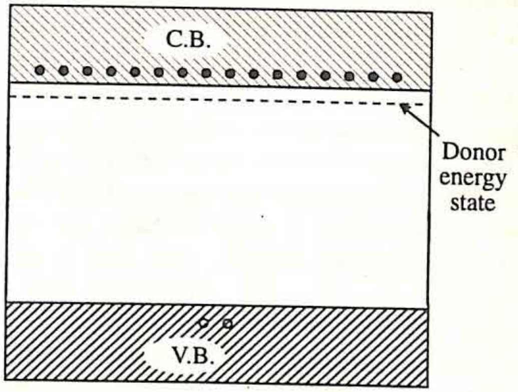


Figure 14.13: Energy level diagram for an n-type semiconductor showing the donor energy levels and electrons in the conduction band at room temperature.

14.3.4 Extrinsic semiconductors

A doped semiconductor or a semiconductor with impurity atoms is called an *extrinsic* semiconductor. If we *dope* silicon, which has four valence electrons, with a controlled amount of pentavalent atoms, say arsenic (A) (or antimony Sb, or phosphorus P), which has five valence electrons, the atoms of the impurity element will substitute the silicon atoms (see Fig. 14.11d). Four of the five valence electrons of As are shared in covalent bonding, while the fifth electron is comparatively free to move as shown in Fig. 14.12. The pentavalent atoms are called the *donor* atoms because they donate electrons to the host crystal and the semiconductor is called *n-type*. On giving up their fifth electron, the donor atoms become positivily charged. However, the material remains electrically neutral as a whole.

The extra electron of the donor atom orbits around the donor nucleus, in a hydrogen like manner. One major difference is that the electron is moving in a dielectric medium, e.g. crystalline silicon or germanium. The former has a relative dielectric constant of nearly ten, so that the donor electron - donor nucleus coulomb attraction is weaker than the electron proton attraction in a hydrogen atom by this factor. You can easily rework, say Bohr's theory of the hydrogen atom for a relative dielectric constant ϵ and find that the energy of the lowest hydrogenic bound state is reduced by a factor ϵ^2. This factor is about a hundred for silicon. Mainly because of this reason, the binding energy of the extra electron, say in phosphorus doped silicon, is 0.045 eV rather than 13.6 eV (hydrogen atom binding energy). Supplying this much energy frees the extra electron. In the band language, we would like that this electron, free to move about, has the lowest possible energy in the empty conduction band. Thus the energy level diagram of a *doped n* type semiconductor is as shown in Fig. 14.13. For phosphorus or arsenic in silicon, the lowest donor electron energy level lies ~ 0.045 eV below the bottom of the conduction band. Note that this energy is comparable to room

temperature thermal energy $k_B(300) \simeq 0.03$ eV, and is much smaller than the energy gap $E_g \simeq 1.1$ eV.

The above fact has a very important consequence. Since it requires much less energy to free a donor electron (or ionize the donor atom) than to promote an electron from the valence band to the conduction band, at any nonzero temperture a sizeable fraction of the donor electrons is in the conduction band because of the favourable Boltzmann factor $\{\exp -(E_g/k_BT)\}$. These electrons can transport electric current. Depending on the temperature and the donor concentration, even for low values of the latter the number of donor generated carriers can be more than the intrinsic carriers. Thus the electrical properties of semiconductors can be modified dramatically, in a controllable way, by adding small amounts of suitable materials. At room temperature, most of the donor atoms are ionized, so that it is a good approximation to assume that the extra donor electrons are all in the conduction band.

The number densities of conduction band electrons n_e and the valence band holes n_h in a doped semiconductor differ from that in a pure semiconductor (the latter being denoted by n_i). It can be shown however, from considerations of thermodynamic equilibrium that

$$n_e n_h = n_i^2. \tag{14.4}$$

This equation is very useful. As already mentioned above, in an n type semiconductor, the donor electrons are almost all free at room temepraturе. Further, in most common cases, this number is much larger than the number of 'intrinsic' conduction band electrons so that one has

$$n_e \simeq N_d >> n_h \tag{14.5}$$

where N_d is the density of donor atoms. For obvious reasons, under these conditions, electrons are called majority carriers and holes are called minority carriers.

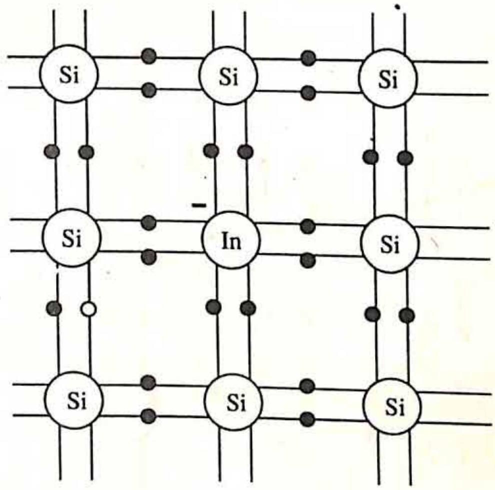


Figure 14.14: Sharing of electrons in covalent bonds with a III group atom present *(p- type semiconductor).*

If we dope intrinsic Si with a controlled amount of trivalent atoms, say indium (In) (or boron B or aluminium Al) which have three valence electrons, impurity atoms will occupy places of some Si atoms and there will be one incomplete covalent bond with a neighbouring Si atom, due to the deficiency of an electron. This is completed by taking an electron from one of the Si-Si bonds, thus completing the In-Si bond. This makes In ionised (negatively charged), and creates a 'hole' or an electron deficiency in the covalently bonded electron system in the crystal as shown in Fig. 14.14. The trivalent atoms are called *acceptor* atoms and the semiconductor is known as *p- type*. The hole is attracted to the negatively charged acceptor nucleus, similar to the attraction between the donor electron and the donor nucleus.

The hydrogen like hole bound state is indicated by an energy level E_B above the top

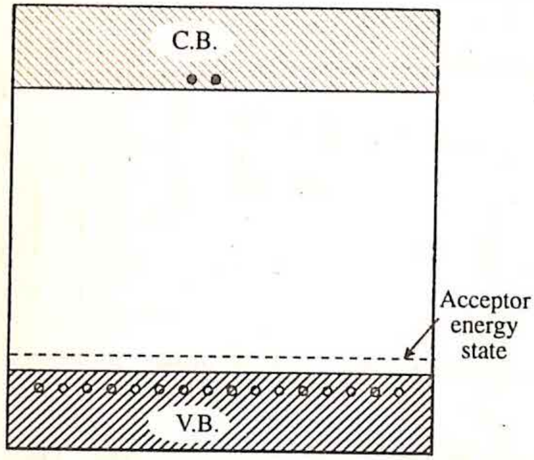


Figure 14.15: Energy band diagram of a p-type semiconductor, showing the acceptor energy levels and holes in the valence band (marked by 0) at room temperature.

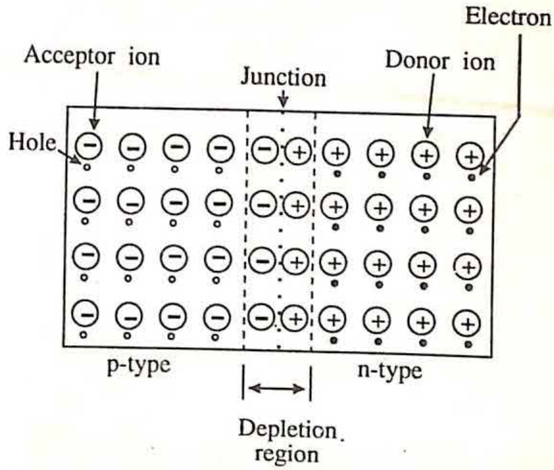


Figure 14.16: Formation of a *pn* junction.

of the valence band (Fig. 14.15). What this means is that on supplying an energy E_B, the hole is free to move around, i.e. there is one electron state missing at the top of the otherwise full valence band. Since $E_B << E_g$, addition of acceptors can modify the electrical properties of semiconductors much the same way as addition of n type impurities (donor).

14.3.5 Semiconductor devices

A p-type or n-type silicon crystal can be grown by adding appropriate impurity in the melt as discussed earlier. These crystals are cut into thin slices called the *wafer*. Semiconductor devices are usually made on these wafers. If on a wafer of n-type silicon, an aluminium film is placed and heated to a high temperature, say 580°C, aluminium diffuses into silicon. In this way a p-type semiconductor is formed on an n-type semiconductor. Such a formation of p-region on n-region is called the *pn junction*. Here, the term junction means the boundary or region of transition between n-type and p-type semiconductor materials. Another way to make a *pn* junction is by diffusion of phosphorus into a p-type semiconductor. There are numerous methods of forming *pn* junctions and it is possible to make more than one junction (e.g. *pnp*) on the same wafer. We shall not go here into the details of these processes.

Such *pn* junctions are used in a host of semiconductor devices of practical applications. The wafer, on which the *pn* junctions are formed, is cut into small pieces. A piece is encapsulated in a casing with electric connections coming out from p and n regions. The simplest of the semiconductor devices is a *pn* junction diode. Another important device is a transistor which has *npn* or *pnp* configuration. We shall learn in the next sections the physics and applications of these devices.

14.3.6 *pn* Junction diode rectifier

Suppose we form a junction of p and n type semiconductors by placing them in contact (e.g. Si doped with Boron and Si doped with phosphorus). Because the p part has too many holes, holes will diffuse to the n

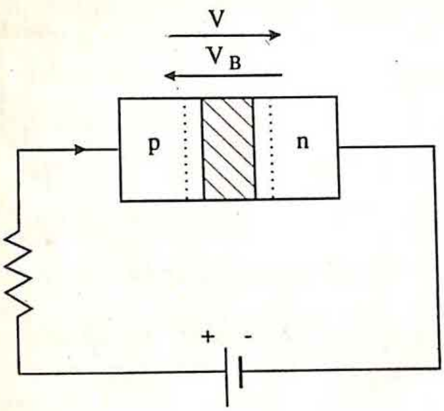


Figure 14.17: Forward biased *pn* junction. The original width of the depletion layer is shown as a dotted line, and the reduced width is shown shaded. The applied electric voltage V across the junction opposes the barrier voltage V_B.

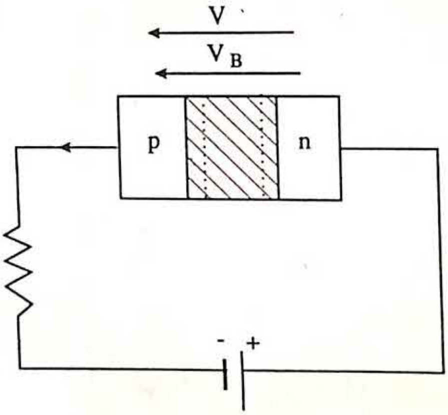


Figure 14.18: Reverse biased *pn* junction. The original depletion layer is shown as a dotted line. On reverse biasing, it becomes thicker, and the applied voltage V is along the barrier voltage V_B.

part, and because the n part has too many electrons, the electrons will diffuse to the p part. The initially unequal concentrations of electrons and holes will result in their diffusion across the junction. This diffusion causes an excess positive charge in the n region (the positive ionic charge not compensated any more by n type carriers) and similarly an excess negative charge in the p region, near the junction (see Fig. 14.16). This double layer of charge creates an electric field which exerts a force on the electrons and holes, against their diffusion. This electric field becomes strong enough as diffusion proceeds, to stop it. In this equilibrium situation, there is a barrier, for charge motion, with the n side at a higher potential than the p side.

The junction region has a very low density of either p or n type carriers, because of interdiffusion. It is called the *depletion region*. There is a barrier V_B associated with it, as described above. Now suppose a dc voltage source is connected across the pn junction. The polarity of this voltage can lead to an electric field across the pn junction that is opposite to that present. The potential drop across the junction decreases, and the diffusion of electrons and holes is thereby increased, resulting in a current in the circuit. This is called *forward biasing*. The depletion region effectively becomes smaller (Fig. 14.17). In the opposite case (Fig. 14.18), called *reverse biasing*, the barrier, increases, the depletion region becomes larger, diffusion is inhibited and the current of electrons and holes is greatly reduced. Thus the *pn* junction allows a much larger current flow in forward biasing than in reverse biasing. This is, crudely, the basis of the action of a pn junction as a rectifier. The actual current voltage characteristics, and the applications are discussed in detail in the next section.

14.4 Semiconductor diode and its applications

The usual semiconductor diode is a single crystal of germanium or silicon. Half of it is doped with n-type impurity. The other half is doped with p-type impurity. Two

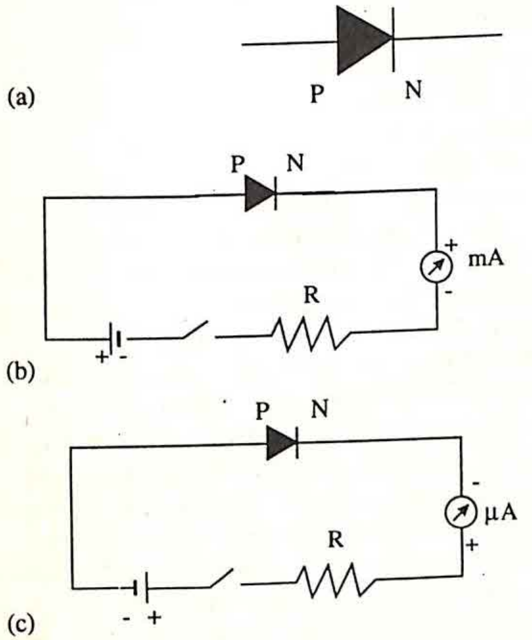


Figure 14.19: (a) Diode symbol, (b) Forward biased diode and (c) Reverse biased diode.

metal wires are introduced into the two regions to act as terminals for electrical connection. Fig. 14.19(a) shows the symbol of a diode. The arrow-head indicates the direction of electrical current. When a battery is connected across the diode such that the positive of the battery is connected to the p-side and the negative to the n-side, current flows in the diode and the diode is said to be forward-biased (Fig. 14.19(b)). If the battery connection is reversed, the diode is said to be reverse-biased (Fig. 14.19(c)) and the current stops.

14.4.1 Characteristics of a pn junction diode

When the diode is forward-biased, the current in the diode changes with the voltage applied across the diode. Fig. 14.20(a) shows the circuit diagram for measuring the diode current for different voltages. The battery is connected to the diode through a potentiometer so that the voltage applied to the diode can be changed. The milliammeter measures the current in the diode and the voltmeter across the diode measures

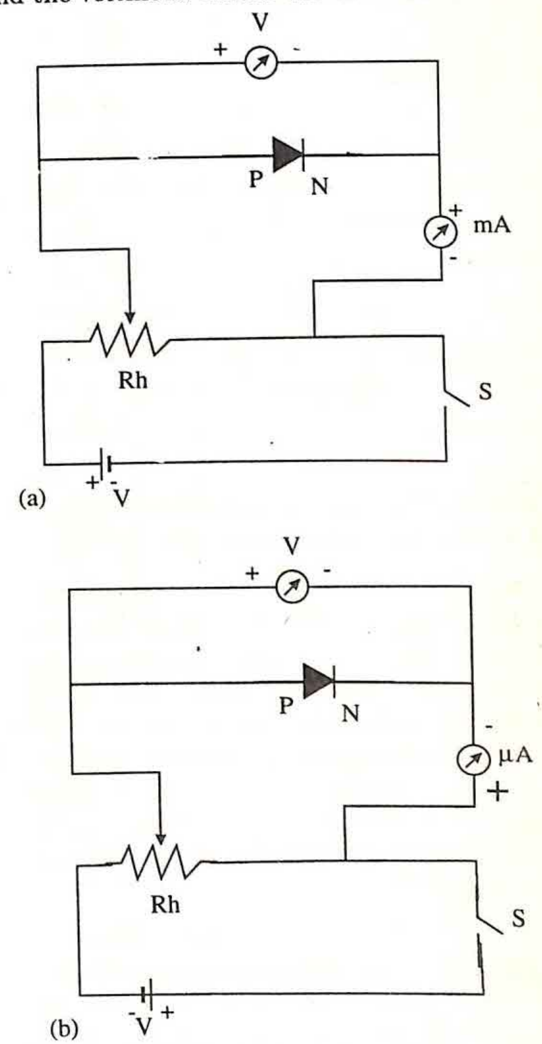


Figure 14.20: (a) Circuit for obtaining the characteristics of a forward biased diode and (b) Circuit for obtaining the characteristics of a reverse bias diode.

the voltage across the diode. The voltage across the diode is increased in small steps of say about 0.1 volt and the current is noted for every voltage. The current increases very slowly till the voltage across the diode

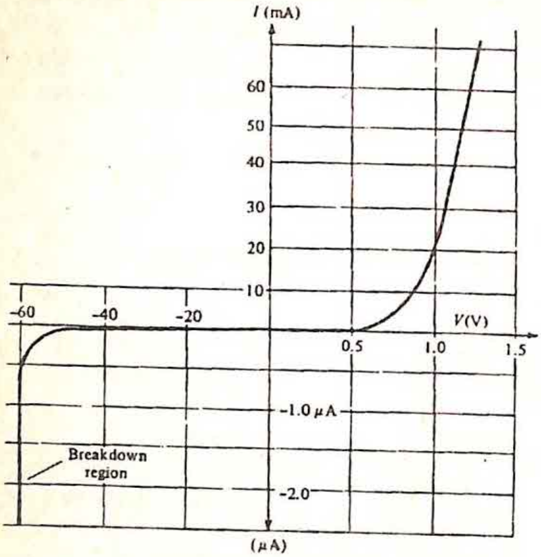


Figure 14.21: Typical diode characteristics in the forward and reverse region (not to scale).

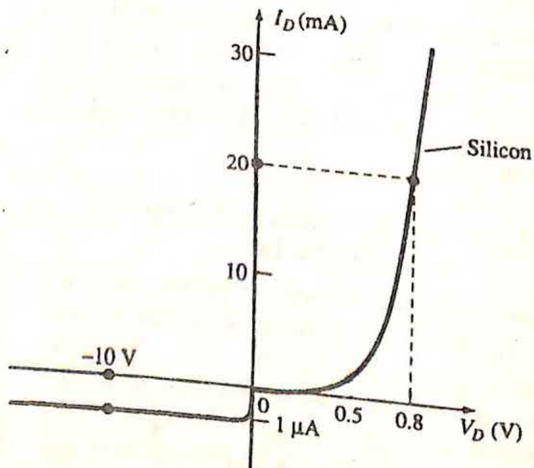
crosses a certain value. After this characteristic voltage, the diode current increases rapidly, even for very small increase in the diode bias voltage. This voltage is called the *threshold voltage or cut-in voltage.* The value of the cut-in voltage is about 0.2V for a germanium diode and 0.7V for a silicon diode. The voltage-current graph is called the forward characteristic and is shown in Fig. 14.21.

When the diode is reverse-biased as in Fig. 14.20(b), the reverse bias voltage produces a very small current, about a few micro amperes which remains nearly constant till a characteristic voltage called the breakdown voltage is reached. Then, the reverse current suddenly increases to a large value. The reverse characteristic of a diode is also shown in Fig. 14.21. The breakdown voltage is also called the peak-inverse voltage of the diode.

A normal pn junction diode is not operated in the reverse bias beyond the breakdown voltage. On the other a special diode called the *Zener diode* is developed, in which the breakdown is profitably utilized to produce constant voltage with useful levels of current. A Zener diode should be operated only in the reverse bias.

A pn junction diode when forward-biased will have a small voltage across it, of the order of the threshold voltage but can give milli-amps to amperes of current. Thus, the diode has a very small resistance in the forward-bias. On the other hand, in the reverse bias, the reverse-bias voltage across the diode can be upto several volts but the current in the diode will be very small, of the order of micro-amperes. Thus, a diode in the reverse-bias has a very large resistance. This property of the diode, namely, a very small resistance in the forward bias and a very large resistance in the reverse bias, means that the diode can conduct well only in one direction. This property is used to convert AC power into DC. Conversion of AC to DC is called rectification.

Example 14.1: The volt-ampere characteristic of a silicon diode is shown in the figure. Determine the resistance of the diode at (a) $I_D = 20$ mA and (b) $V_D = -10$V.



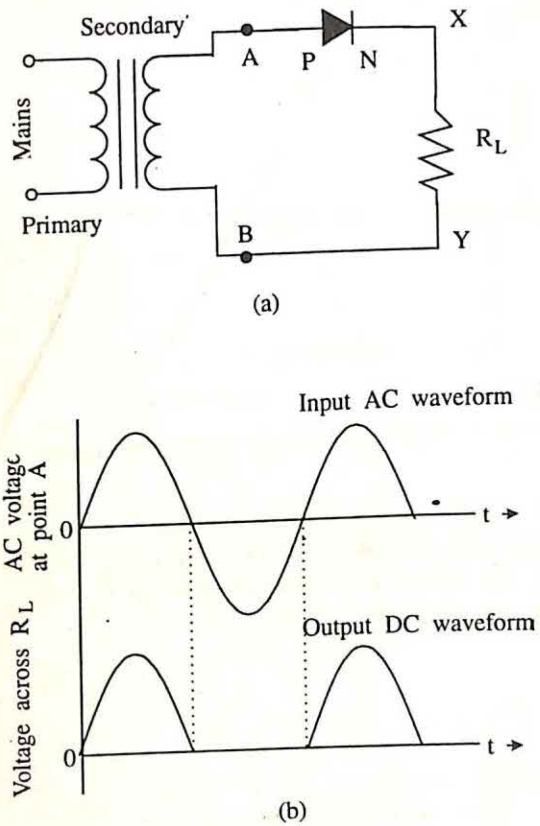


Figure 14.22: (a) Half wave rectifier and (b) AC and DC voltage waveforms in a half wave rectifier.

Answer: Considering the diode characteristics as a straight line passing through the origin, we can calculate the resistance using Ohm's law.

(a) From the curve, at $I_D = 20\text{mA}$, $V_D = 0.8\text{V}$ and

$$R_D = \frac{V_D}{I_D} = \frac{0.8\text{V}}{20\text{mA}} = 40\Omega$$

(b) From the curve, at $V_D = -10\text{V}$, $I_D = -1\mu\text{A}$ and

$$R_D = \frac{V_D}{I_D} = \frac{10\text{V}}{1\mu\text{A}} = 10\text{M}\Omega$$

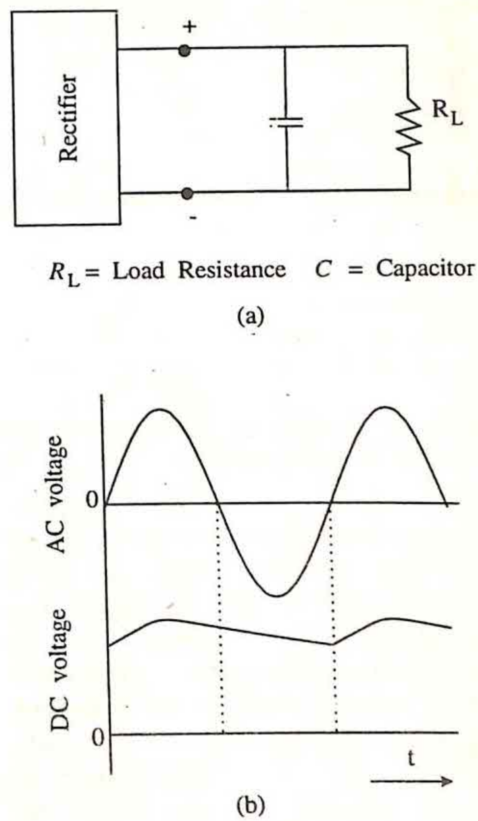


Figure 14.23: (a) Rectifier with a capacitor filter and (b) AC and DC voltage waveforms in a rectifier with a capacitor filter.

This shows that the diode has small resistance in the forward bias and a very large resistance in the reverse bias.

14.4.2 Diode as a rectifier

A simple rectifier circuit called the half wave rectifier, using only one diode, is shown in Fig. 14.22(a). The transformer converts the mains voltage to the required value in the secondary.

When the voltage at A is positive, the diode conducts and when the voltage at A is negative, the diode is reverse-biased and

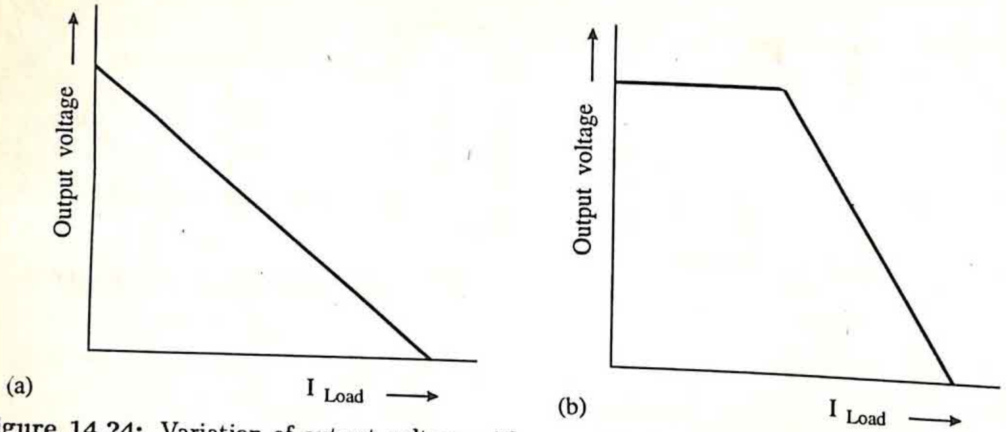
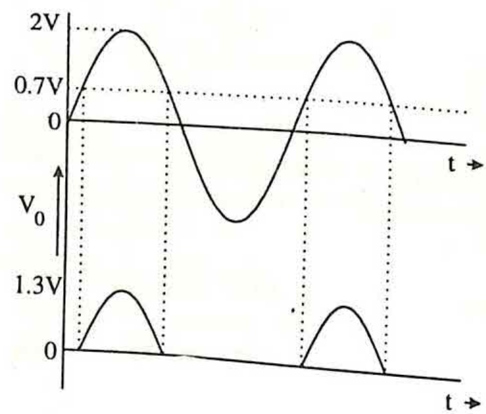


Figure 14.24: Variation of output voltage with current in load for (a) unregulated power supply and (b) regulated power supply.

does not conduct. So, the AC voltage at point A and the voltage across R_L are as shown in Fig. 14.22(b). Since the diode conducts only in the positive half-cycles, the voltage between X and Y across R_L will be DC but will be in pulses. When this is given to a circuit called the filter, it will smoothen the pulses and will produce a rather steady DC voltage. There are different types of filters. The simplest filter is a large electrolytic capacitor. This is often adequate for low voltage rectifiers. A rectifier circuit connected with a Capacitor-filter is shown in Fig. 14.23(a). When a resistance is connected across the filter, a current flows in it and the output voltage decreases a little. When the resistance value is decreased, i.e., when more current is drawn, the output voltage decrease further. However, with the help of additional circuit, the output voltage can be kept constant even when current is drawn. Such a power supply is called a *regulated power supply*. The variation of output voltages with load for unregulated and regulated power supplies are shown in Fig. 14.24.

Example 14.2 : Consider the half wave rectifier shown in Fig. 14.22(a). Assume the diode to be a silicon diode with a threshold voltage of 0.7V. Draw the output waveform if the input is a sine wave with an amplitude of 2V.



Answer : Since the diode threshold voltage is 0.7V, we can assume that the diode does not conduct till the input voltage reaches 0.7V. So, it will conduct from 0.7V to 2V and again till the input voltage falls down to 0.7V. So, the input and output waveform for the half wave rectifier will be as shown in the figure.

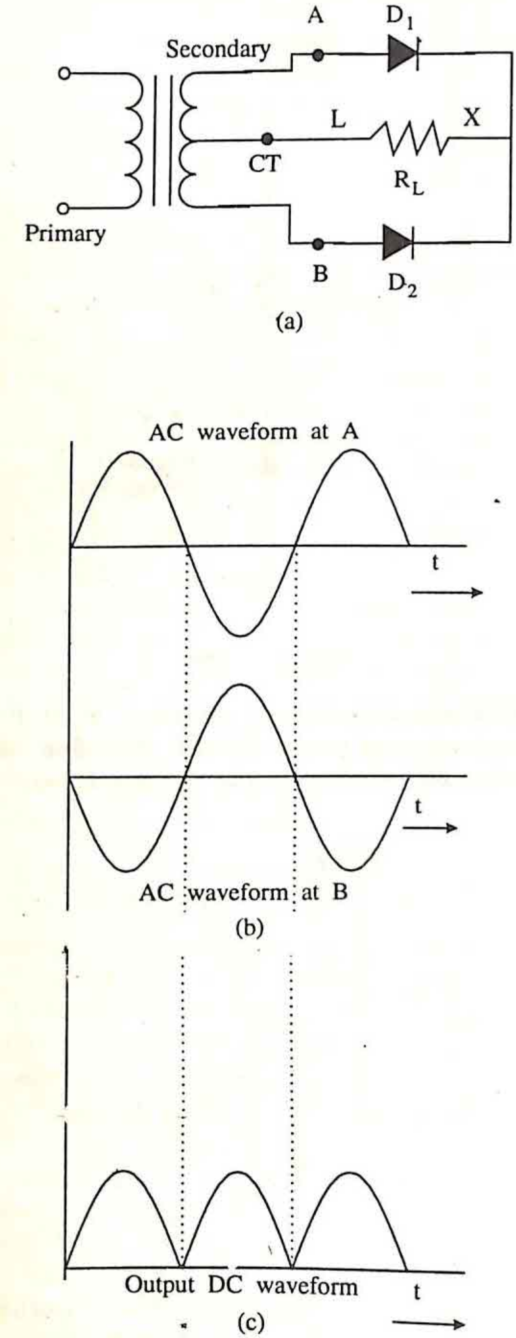


Figure 14.25: (a) Full wave rectifier; (b) AC voltage waveforms at points A and B, and (c) Output DC waveforms of a full wave rectifier.

14.4.3 Full wave rectifier

Fig. 14.25 shows a circuit which is called a full-wave rectifier. It uses two diodes. The secondary of the transformer is wound in two parts and the junction is called a centre-tap. The AC voltage at A and B with respect to the centre-tap are out of phase with each other, as shown in Fig. 15.25(b). Therefore, diode D_1 conducts during the positive half-cycle of AC voltage at A and diode D_2 conducts during the positive half-cycle of the AC voltage at B. So, the DC output voltage of a full-wave rectifier is two times that of the half-wave rectifier, as shown in Fig. 14.25(c).

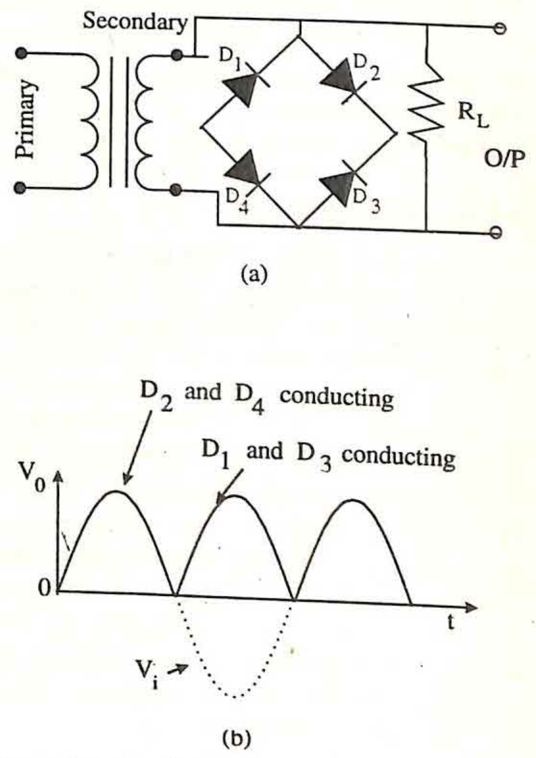


Figure 14.26: (a) Bridge rectifier and (b) Input and output waveforms for a bridge rectifier.

Another rectifier circuit called the bridge rectifier which uses four diodes is shown in Fig. 14.26(a). Here the secondary of the

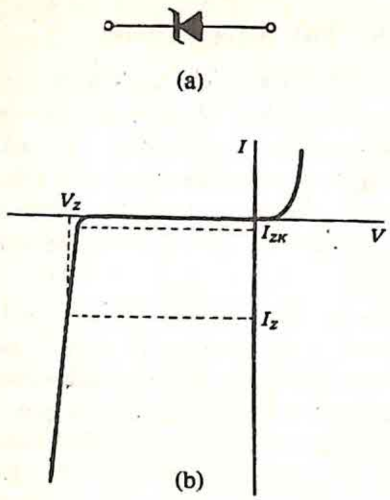


Figure 14.27: (a) Symbol for a zener diode and (b) Typical characteristics of a zener diode.

transformer has no centre-tap. Thus one secondary is adequate. The AC and DC wave forms are also shown in Fig. 14.26(b).

14.4.4 Zener diode

A zener diode works in the reverse bias and when the reverse bias is equal to the breakdown voltage, the voltage across the zener remains almost constant and the current increases rapidly. A zener characteristic is shown in Fig. 14.27. A zener diode is often used to produce a constant voltage. Fig. 14.28. shows how an unregulated DC voltage V_i can be regulated to give a constant output voltage $V_0 (= V_Z)$ using a zener diode. The voltage across the zener diode will be almost constant, equal to its breakdown voltage V_Z. If V_i changes, correspondingly V_R changes since V_Z remains constant. Thus, $V_0 = V_Z$ and remains constant.

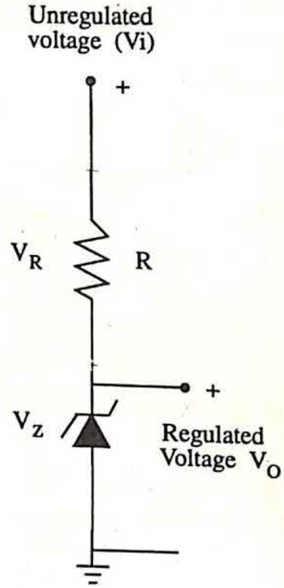
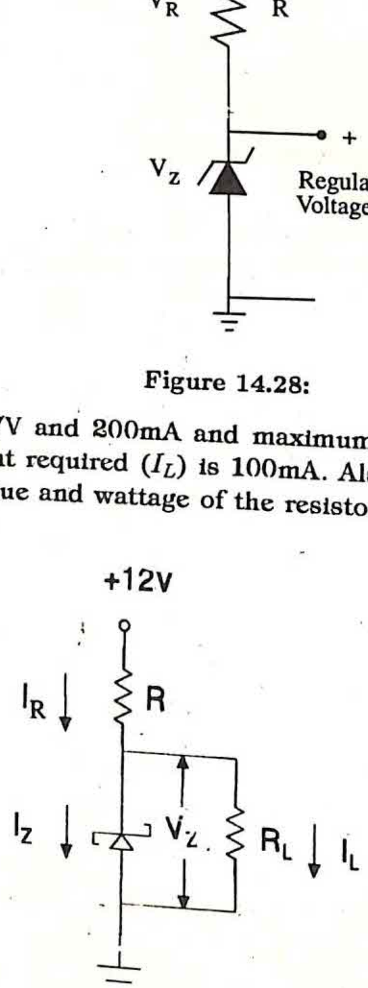


Figure 14.28:

Example 14.3: A zener diode is connected in the circuit as shown in the figure. What is the voltage drop across the resistor R and current through it, if the voltage and current specifications of the zener diode are 4.7V and 200mA and maximum load current required (I_L) is 100mA. Also find the value and wattage of the resistor R used.



Answer: The break down voltage V_z of the zener diode is 4.7V. The voltage distribution between R and zener is as shown in the figure.

From the figure we can easily see that V_R

$= 12V - V_z = 7.3V.$

From the figure we can also see that the current I_R is getting divided into two parts viz., I_R (current through zener) and I_L (current through the load resistor). When there is no load connected, the total current I_R has to pass through the zener.

So, the maximum I_R should be less than the maximum current the zener can take in order to operate the zener diode safely. Also, the zener diode requires some minimum current to pass through it for proper operation. Let us assume that it is 10mA. So the current in R should be atleast 10mA more than the maximum current required to go through the load resistor R_L. The maximum current required to go through R_L is 100mA. Therefore, I_R should be atleast 110mA.

We can keep $I_R = 150$mA to ensure good working (The given zener diode can take a maximum of 200mA). The value of the resistor R needed to allow 150mA current (for a voltage of 7.3V across R) is,

$$R = V_R/I_R = 7.3V/150mA$$
$$= 7.3V/0.15A = 48.66\Omega$$

We can use 47 Ω or 50 Ω resistors which are readily available in the market.

Wattage is another important specification one has to consider while choosing a component. When some current passes through a component, it generates heat because of the power dissipated by the component. This heat is to be dissipated properly to ensure proper operation. Components with different ranges of heat dissipation capabilities are available in the market. To choose the correct component, we have to find out the wattage of the component to be used.

$$\text{The wattage of } R = V_R \times I_R$$
$$= 7.3V \times 150mA$$
$$= 7.3V \times 0.15A$$
$$= 1.095W$$

So, it is safe to use a 2W resistor.

14.4.5 Regulated power supplies

We can convert AC into DC by rectification and filtering. As mentioned above, when we connect a resistance at the output of the rectifier, the DC voltage value drops. Similarly if the AC voltage at the input is changed, the output DC value changes. In real applications, we generally want the DC voltage to be constant, even when we draw current from it or when the input AC voltage changes. Such constant voltage power supplies are called *regulated power supplies*. Nowadays, we get such regulated power supplies in the form of integrated circuits. We can connect the output from the filter to the voltage regulator IC as prescribed by the manufacturer. One such voltage regulator IC is IC3085A. A power supply circuit using IC3085 is shown in Fig. 14.29. With this IC3085A, we can vary the output voltage from 2V - 37V approximately, by varying the potentiometer.

Example 14.4: In the potential divider circuit of a IC3085A based voltage regulator, the resistors used are $R_1 = 85K$ and $R_2 = 15K$. If the output voltage is 10V, find the reference voltage.

Answer: The voltage regulator with potential divider is as shown in the figure below. In voltage regulator IC, a constant DC voltage called the reference voltage V_{ref} is provided (built in) and the voltage across R_2 (i.e. $VoR_2/(R_1 + R_2)$) is compared with the V_{ref}. When the voltage across R_2 is equal to the reference voltage, the output remains constant.

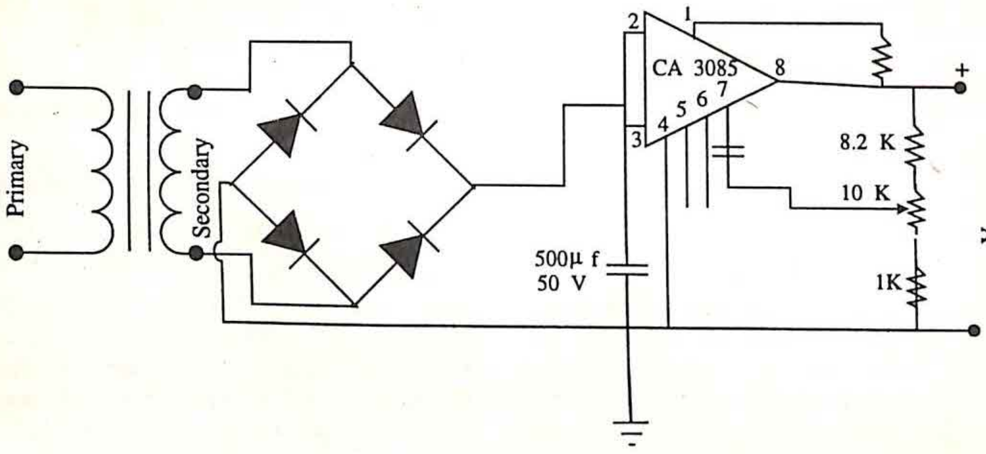
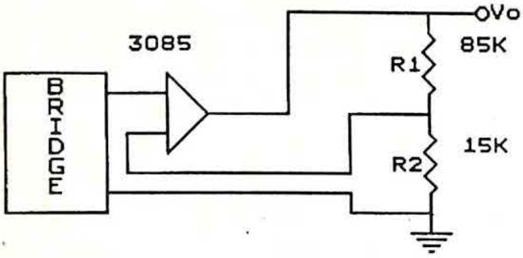


Figure 14.29: A power regulator circuit, using IC3085A.



As shown in the figure, the output is divided using the potential divider circuit and compared with the internal constant reference voltage. The voltage V_0 is divided between the 85K and 15K resistors in 85 : 15 ratio. So, the voltage drop across the 15K resistor can be obtained by using the formula,

$$V_{\text{ref}} = \frac{V_0 R_2}{R_1 + R_2}$$
$$= \frac{10\text{V} \times 15\text{K}}{85\text{K} + 15\text{K}} = 1.5\text{V}$$

Another kind of IC voltage regulator is called the '3-terminal' regulator. It is a simple IC with three pins. It can be connected as shown in the Fig 14.30 to obtain a regulated output. The IC7805 shown in Fig. 14.30, gives a constant output voltage of 5V.]

Thus integrated circuits are very convenient and easy to use. The manufactures will tell us how to connect them in a circuit to obtain the required function. We can combine such IC's to make bigger total instruments. Nowadays, this 'system approach' using IC's is a common practice in electronics.

Example 14.5: A 7805 (3-pin voltage regulator) (Fig. 14.30) based power supply is designed to give 5V output with a maximum load current of 500 mA. Find the maximum value of the load that the power supply can drive.

Answer: As given in the problem, the '7805' based voltage regulator circuit can deliver a maximum of 500 mA. The load resistance to take the maximum current is

$$R_{L(max)} = \frac{V_{\text{out}}}{I_{\max}}$$
$$= \frac{5\text{V}}{500\text{mA}} = \frac{5\text{V}}{0.5\text{A}}$$
$$= 10\Omega.$$

And the Wattage of the resistor is $V \times I_0 = 5\text{V} \times 500\text{mA} = 2.5$ W

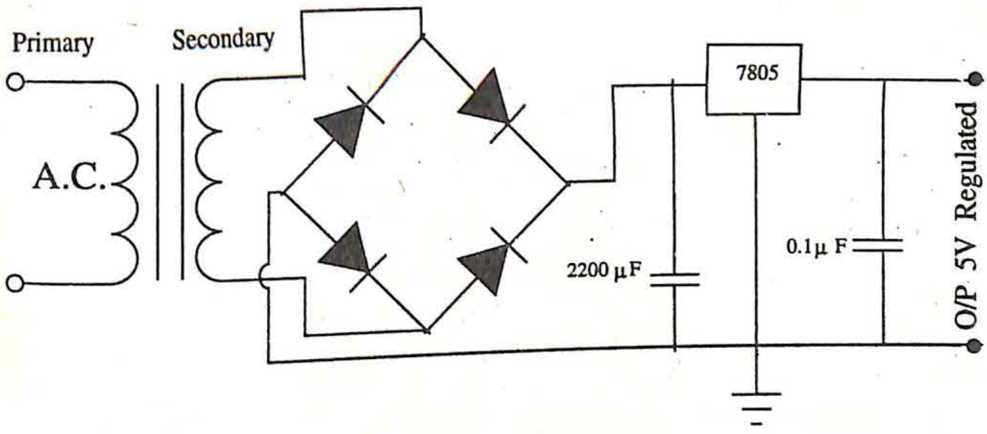
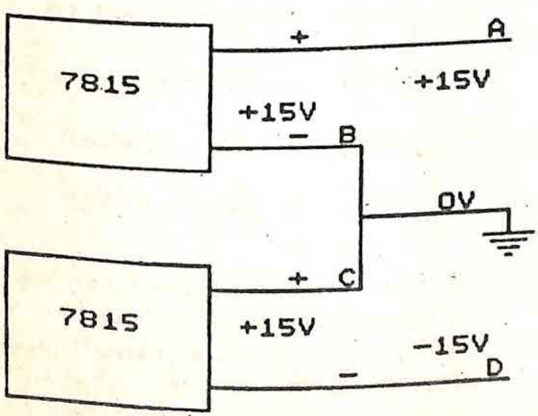


Figure 14.30: A '3-terminal' regulator.

Even though the power supply can supply a maximum current of 500 mA, it is advisable to use it with the loads which need only 50 to 75% of the maximum current for safety and long life.

So the maximum permissible load current is 250 to 375 mA and the load permissible is 5 V/375mA = 5V/0.375A = 13.33 Ω.

Example 14.6: How can you make a ±15V power supply using two 7815 IC's (3-pin voltage regulators)?



Answer: Let us take two power supplies using 7815's and connect them as shown in the figure. The terminal B (GND) of power supply I is connected to the terminal 'C' of the power supply II. Now with respect to B and C,A is at +15 V potential and D is at −15 V potential.

Operational amplifiers require usually ± 15 V power supply. Three pin voltage regulators are available for different fixed voltages viz. 7806(+6 V), 7812(+12 V), 7815(+15 V), etc.,. It is very convenient to connect two 7815 based postive voltage regulators in the above mentioned way to make ± 15 V power supply.

14.5 Transistor

14.5.1 Introduction and characteristics

A transistor can be considered as a thin wafer of one type of semiconductor between two layers of another type. The *npn* transistor has a *p*-type wafer between two *n*-type layers. Similarly the *pnp* transistor has a *n*-type wafer between two *p*-type layers. Fig.

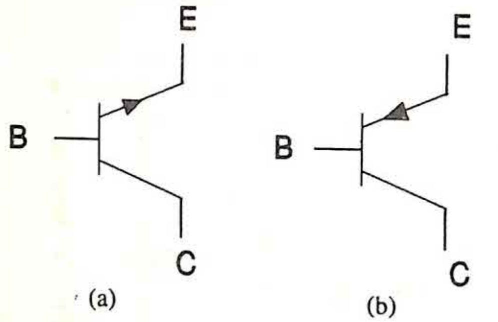


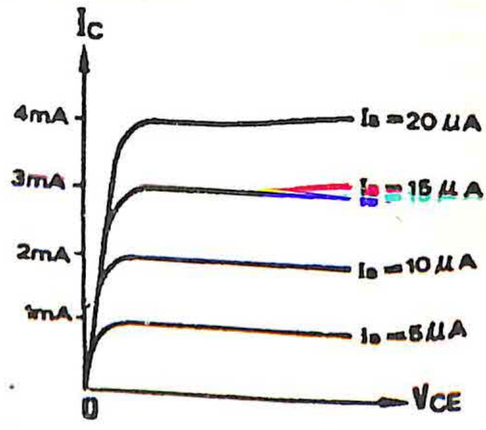
Figure 14.31: Symbol for (a) npn, and (b) pnp transistors.

14.31 shows the symbols of two types of transistors. The three portions of the transistor are called emitter (E), base (B) and collector(C) as shown in the figure. If we apply suitable voltages across the three terminals as shown in Fig. 14.32, currents flow in the input and output circuits. The currents flowing through the transistor are (i) the emitter current I_E; (ii) base current I_B and (iii) collector current I_C. When the currents are flowing into the transistor, they are taken as positive, by convention. The most versatile and often used arrangement is the common emitter (CE) circuit and we will now learn about it briefly. Let us concentrate on *npn* transistor. The same will hold good for *pnp* transistors also.

To measure the currents flowing in the transistor, we can connect the circuit as shown in Fig. 14.32. Two types of measurements can be made. One is to obtain the input characteristics and the other is to obtain the output characteristics. To measure the input characteristics, the voltage V_{CE} between the collector and emitter is adjusted to have a suitable value. Then $V_{\rm BE}$ is increased from zero in small steps and in each case I_B is noted. A graph showing the variation of I_B with $V_{\rm BE}$ is drawn. The experiment is repeated for other values of V_{CE}. Thus, a family of curves is obtained called the input characteristics (Fig. 14.33(a)).

To measure the output characteristics, $V_{\rm BE}$ is adjusted to have a suitable value so that I_B will have a small value of a few micro amperes. For different values of V_{CE}, the corresponding values of I_C are noted. A graph of I_C and V_{CE} is drawn. The experiment is repeated for different values of I_B and in each case a graph showing I_C versus V_{CE} is plotted. $I_C - V_{CE}$ curves constitute a family of output characteristics as shown in Fig. 14.33(b). From Fig. 14.33(a) it is clear that input characteristics appear just like the diode characteristics. The output characteristics show that I_C changes rapidly in the beginning but soon I_C is more or less independent of V_{CE}. The ratio of I_C/I_B is nearly constant and is called the current gain, denoted by β.

Example 14.7: The output characteristics of a transistor is common emitter configuration are shown in the figure. Find the current gain.



Answer: The output charactistic is measured by fixing the base voltage V_{BE} to give a few micro amperes of base current, and measuring the collector current as a function of V_{CE}. The output characteristic curves are

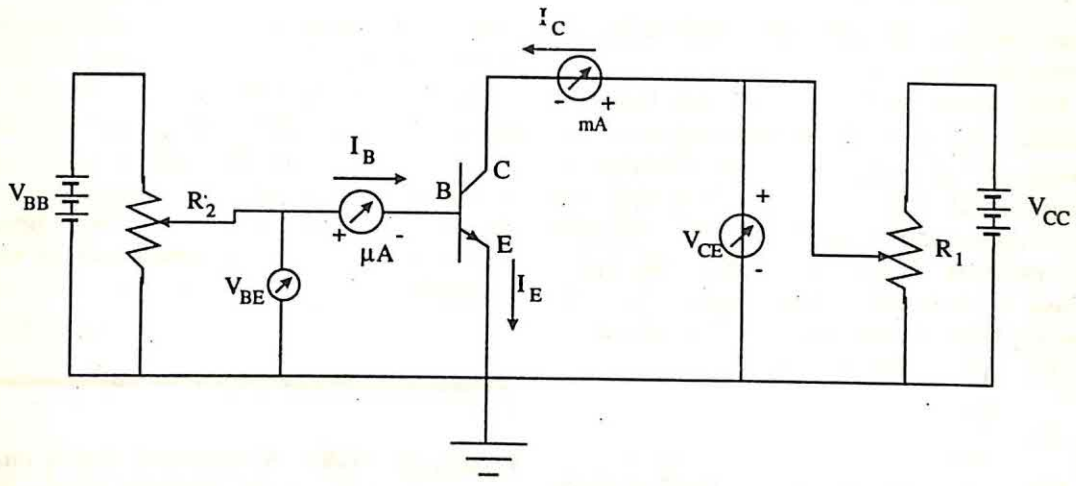
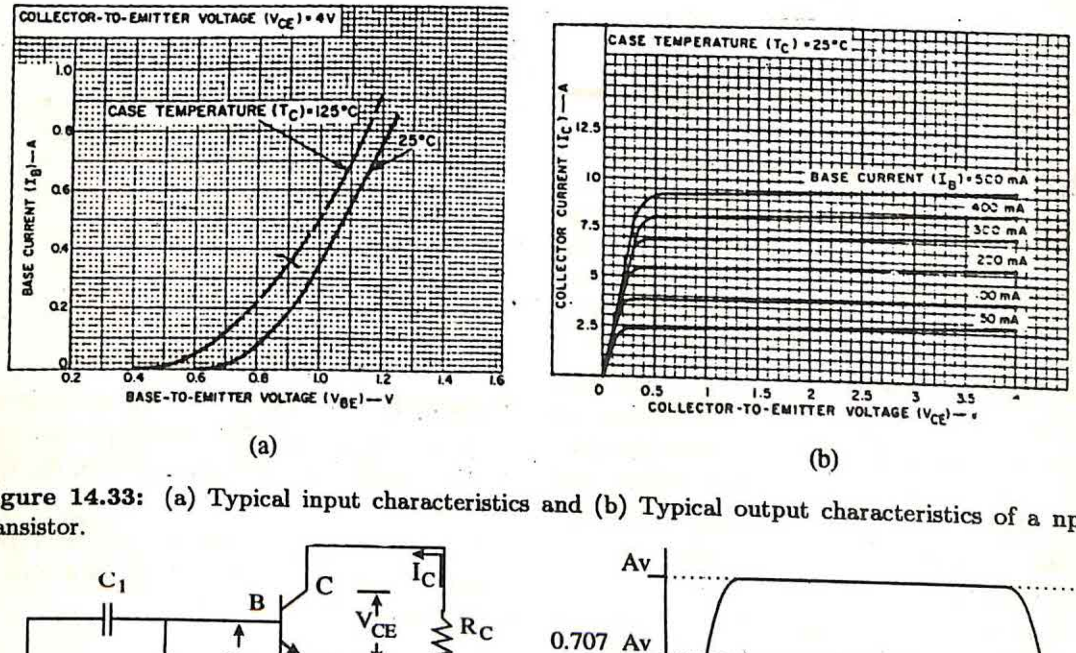


Figure 14.32: Circuit for obtaining the characteristics of a npn transistor.



(a) (b)

Figure 14.33: (a) Typical input characteristics and (b) Typical output characteristics of a npn transistor.

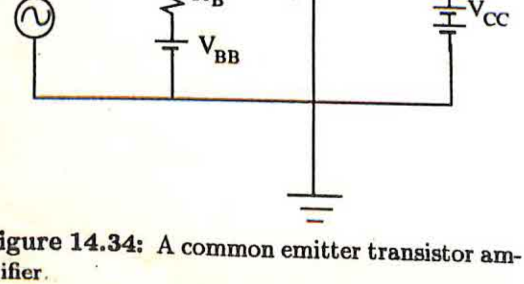


Figure 14.34: A common emitter transistor amplifier.

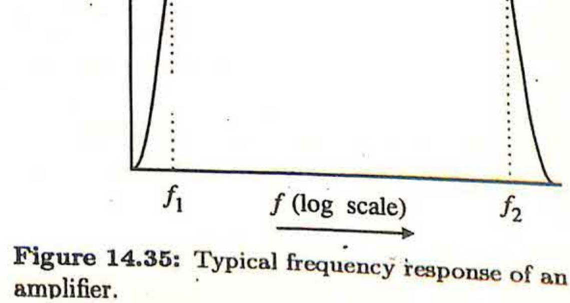


Figure 14.35: Typical frequency response of an amplifier.

thus obtained for different values of I_B, as given in the figure.

From the output characteristics, thus obtained, we can find the ratio between the change in the collector current to change in base current. Initially, for change in V_{CE}, the I_C changes rapidly and later it will be almost independent of V_{CE}. We work in this region where I_C is nearly independent of V_{CE}. So we can take the current gain β as change in I_C for a given change in I_B.

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$

From the figure, for $\Delta I_C = 4$ mA $-$ 2mA, $\Delta I_B = 20\mu\text{A} - 10\mu\text{A}$

$$\beta = \frac{4\text{mA} - 2\text{mA}}{20\mu\text{A} - 10\mu\text{A}}$$

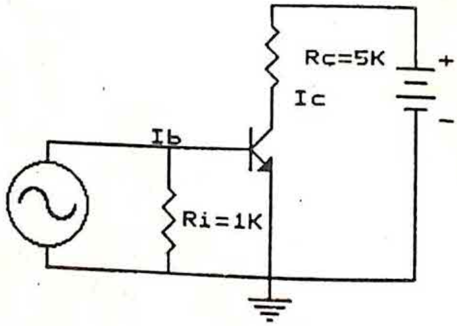
$$= \frac{2 \times 10^{-3}}{10 \times 10^{-6}} = 200.$$

14.5.2 Transistor as an amplifier

Fig. 14.34 shows an *npn* transistor circuit as an amplifier. The transistor is used in the common emitter arrangement. The AC signal is applied at the base and the output is taken at the collector. In the absence of the AC signal, batteries V_{BB}, V_{CC}, resistances R_C and R_B provide suitable voltages to the base, emitter and collector. The base emitter junction should be forward-biased and the collector base junction should be reverse-biased. Under this condition, the transistor will work as an amplifier.

The AC signal V_s is amplified to a value V_0 at the output. The ratio V_0/V_s is called the gain of the amplifier. When the frequency of the input signal is increased from a small value, the value of the gain increases initially and remains constant upto a certain frequency and decreases thereafter with further increase of frequency, as shown in Fig. 14.34. The constant value of the gain is called the gain of the amplifier and the difference between the frequencies f_1 and f_2 is called the band-width. At f_1 and f_2, the gain falls to $1/\sqrt{2}$ of the constant value. It means that the amplifier can amplify input signals with a gain of A if the input signal frequency lies within the bandwidth of the amplifier.

Example 14.8: A transistor has a current gain of 50. If the collector resistance is 5KΩ and the input resistance is 1KΩ (approx.), find the voltage gain of the amplifier.



Answer: The transistor has an internal resistance of about 1KΩ between base and emitter which is shown pictorially in the figure above.

The ratio of output voltage V_0 to the input voltage V_i is called the voltage gain and is represented by the formula:

$$A_v = V_0/V_i \qquad (i)$$

V_0 can be written as a product of the collector current and collector resistance.

$$\text{i.e.} \quad V_0 = I_c R_c \qquad (ii)$$

Similarly V_i is the product of the input current and input resistance.

$$\text{i.e.}\quad V_i = I_B R_i \qquad (iii)$$

By substituting Eq. (ii) and (iii) in (i) we have:

$$A_v = (I_C/I_B)(R_C/R_i)$$

But we know that $I_C/I_B = \beta$ (current gain). So, we can write A_v as a product of β and R_C/R_B.

$$A_v = \beta R_C/R_B = 50 \times 5K/1K = 250$$

(But in practice, we may achieve a much less gain due to capacitors and other wiring effects which we will not worry about here).

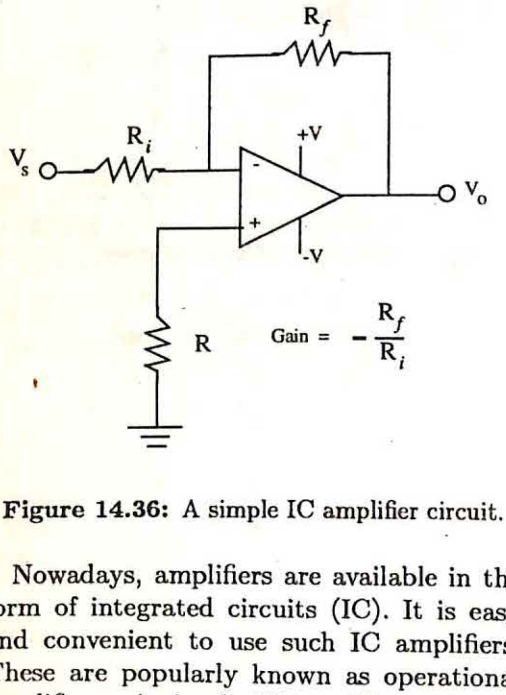


Figure 14.36: A simple IC amplifier circuit.

Nowadays, amplifiers are available in the form of integrated circuits (IC). It is easy and convenient to use such IC amplifiers. These are popularly known as operational amplifiers. A simple IC amplifier circuit is shown in Fig. 14.36. We can see that the gain of the amplifier is given by the values of the two resistors connected by us. So by connecting a power supply and two resistors, we get the amplifier to work.

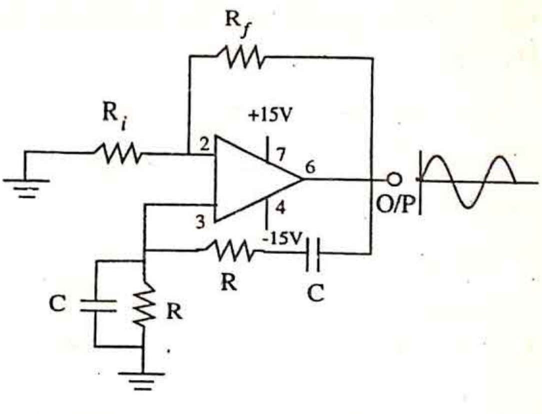


Figure 14.37: Circuit diagram of a Wien-bridge oscillator.

14.5.3 Oscillators

Another circuit which is commonly used is called the oscillator. It produces an AC output without any input. These oscillators can be built using amplifiers and by connecting the output to the input in a suitable way. Since amplifiers are available as IC's, it is easy to make oscillators using IC amplifiers. Fig. 14.37 shows the circuit of an oscillator using an IC amplifier. The frequency of the AC output from the oscillator is given by $1/(2\pi RC)$. Thus, by changing the value of R and C, we can change the frequency of the oscillator (We can also build oscillators using a combination of inductance and capacitance instead of resistance and capacitance).

Example 14.9: In an operational amplifier 741 base Wien Bridge oscillator circuit as shown in Fig. 14.37 R=1K, C=0.22μF, R_i= 10K and **R_F = 30K. Find the gain and frequency of the oscillator.

Answer: For Wien Bridge oscillator to work well, the gain of the amplifier should be about 3. We know that the gain is given

by R_f/R_i.R_f = 30K and R_i = 10K. Hence the gain is:

$$A_v = -R_f/R_i = -30\text{K}/10\text{K} = -3.$$

The frequency of the AC output of the oscillator is given by

$$\nu = 1/(2\pi RC)$$

Subsittuting the values of R = 1K and $C = 0.22\mu$F in the above formula to get:

$$\begin{aligned}\nu &= 1/(2(22/7)1\text{K} \times 0.22\mu\text{F}) \\ &= 7/(44 \times 1000 \times 2.2 \times 10^{-7})\text{Hz} \\ &= 723.14\text{Hz}\end{aligned}$$

14.6 Digital electronics

14.6.1 Introduction

We have already learnt about number systems - like binary and decimal systems. We also know how to convert a number from one system to another. In the binary number system, as you know, we have only two digits 0,1 (decimal system has ten digits). A pulse waveform (Fig. 14.38) has only two voltage levels. By using these two levels to represent the binary digits 0 and 1, we can develop a new subject - Digital Electronics. This section is intended to provide the first step in our understanding of digital electronics. We will now learn about some of the basic building blocks of digital electronics like logic gates.

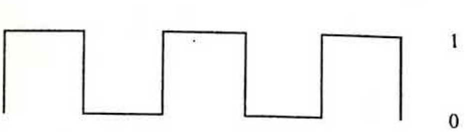
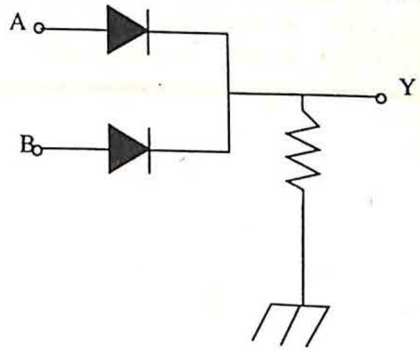


Figure 14.38: A digital signal.

14.6.2 Logic gates

Logic gates are important building blocks in digital electronics. These are circuits with one or more inputs and one output. The basic gates are OR, AND, NAND, and NOR. As you know, in digital electronics, only two voltage levels are present. Conventionally these are 5V and 0V, referred to as 1 and 0 respectively or vice versa, depending on the convention adopted. They are also referred to as 'high' (5V) and 'low' (0V) or vice versa. We will refer to 5V as (1) and 0V as (0).



(a) OR Gate - Circuit diagram

Truth Table for OR Gate

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(b)

Figure 14.39: (a) Circuit diagram and (b) Truth Table for a OR gate.

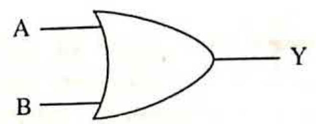
(i) OR Gate

The circuit shown in Fig. 14.39(a) is a OR gate. The output Y will be 1 (ie. 5V) when the A input is 1 OR when the B input is

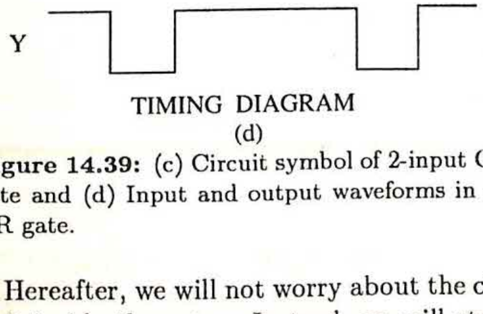
1 OR when both are 1. This is written as $Y = A + B$.

The + sign is read as OR. The table (14.39(b)) giving the different possible inputs and the corresponding outputs is called the "Truth Table". From the "Truth Table" we can see that the output is 1 when A OR B has an input of 1.

Using such gates, bigger circuits and total systems can be built. Then, it is inconvenient or cumbersome to draw the complete circuit diagrams using gates showing diodes, transistors etc,. Thus, symbols are used to represent the gates. For example, the symbol of a two input OR gate is represented as shown in Fig. 14.39(c). A and B are inputs and Y is the output. The timing diagram (Fig. 14.39(d)) shows how the output waveform changes for given input waveform.



CIRCUIT SYMBOL OF 2-INPUT OR GATE
(c)

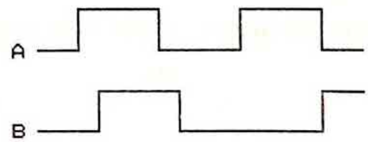


TIMING DIAGRAM
(d)

Figure 14.39: (c) Circuit symbol of 2-input OR gate and (d) Input and output waveforms in an OR gate.

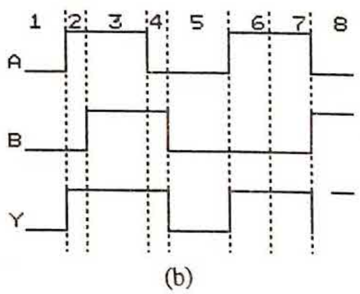
Hereafter, we will not worry about the circuit inside the gates. Instead, we will start with gate as the basic unit and learn its functions and properties. We can later learn to use these building blocks to build circuits for specific purposes.

Example 14.10: Write the output waveform of the OR gate for the given inputs (Fig. (a)).



(a)

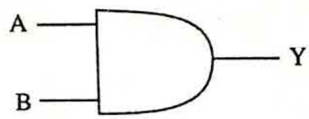
Answer: The output of the OR gate is high when either A or B or both the inputs are high. If both the inputs are low, the output is low.



(b)

Consider the input wave forms. During the period 1, both A and B are zero and the output Y is therefore zero. During the time interval 2, $A = 1$ and $B = 0$, and hence the output is high. We can continue to examine A and B waveforms in a similar way to work out the output waveform. However, observe the interval 7 carefully. Both A and B are changing the states (A from high to low and B from low to high). At this juncture, there is a possibility for both A and B being low

momentarily. So, the output is low for a very short duration and goes high. This is not a desirable situation and should be avoided in the design. The input - output waveforms are shown in Fig. (b).

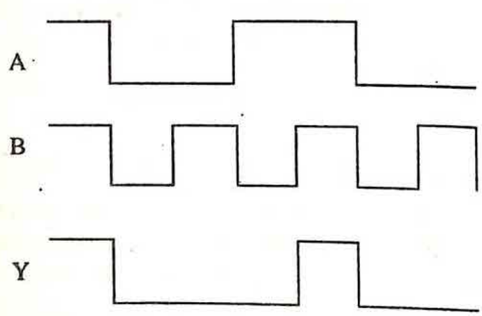


CIRCUIT SYMBOL OF 2-INPUT AND GATE
(a)

TRUTH TABLE OF 2 INPUT AND GATE

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)



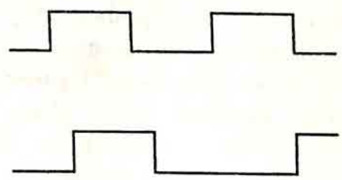
TIMING DIAGRAM
(c)

Figure 14.40: (a) Circuit symbol, (b) Truth table and (c) Timing diagram of an AND gate.

(ii) AND gate

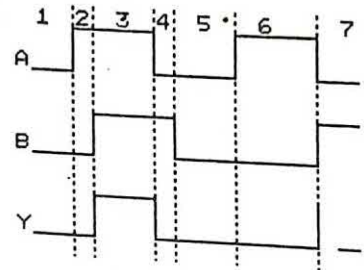
The symbol of a simple AND gate with two inputs A and B and output Y is shown in Fig. 14.40 along with the Truth Table. This is written as $Y = AB$.

Example 14.11: Two input waveforms A and B shown in Fig. (a) are applied to an AND gate. Write the output waveform.



(a)

Answer: The output of an AND gate is high only when both the inputs A and B are high. Otherwise, the output is low. Using this property and analysing on the same lines as in example 14.10, we get the following waveform at the output (Fig. (b)).



(b)

(iii) NOT gate

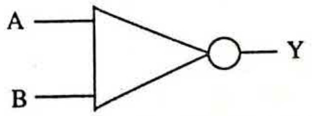
This has one input and one output. The output is the inverse of the input. When A is 1, the output will be 0 and vice versa. The symbol and the Truth Table of NOT gate

are shown in Fig. 14.41. This is written as $Y = \bar{A}$..

(iv) NAND gate

We have just learned the function of the NOT gate. A NAND gate can be obtained by combining AND and NOT gates as shown in Fig. 14.42. Inverse of AND output is NAND. This is written as $Y = \overline{AB}$

Truth table for a two input NAND gate and the NAND gate symbol are shown in the figure.



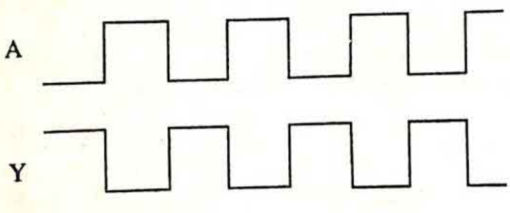
CIRCUIT SYMBOL OF NOT GATE

(a)

TRUTH TABLE OF NOT GATE

A	Y
0	1
1	0

(b)



TIMING DIAGRAM

(c)

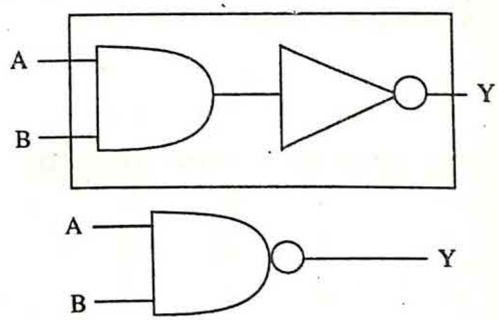
Figure 14.41: (a) Circuit symbol, (b) Truth table and (c) Timing diagram of a NOT gate.

(v) NOR gate

NOR can be realized using OR and NOT gates as shown in Fig. 14.43. The truth table and the circuit symbol are as shown. This is written as $Y = \overline{A + B}$.

(vi) Exclusive-OR (X-OR) gate

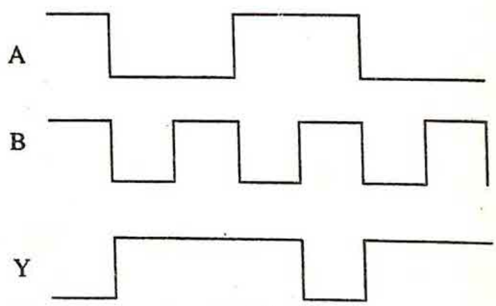
An X-OR gate is a special gate which can be realized by using NOT, AND and OR gates. It gives a high output (i.e. 1) when an odd number of inputs is high. (The output has the value 1 if either A or B has value 1, but not both; if both are 0 or both are 1, the output is 0). The Truth table and circuit symbol for a two input X-OR gate are given in Fig. 14.44. This is written as $Y = A \oplus B$.



(a)

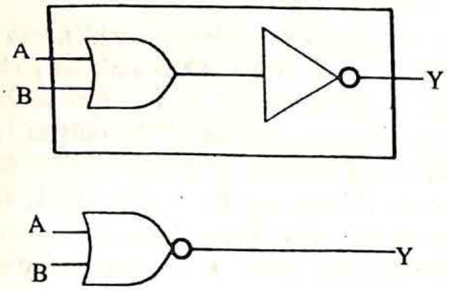
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

(b)



(c)

Figure 14.42: (a) Circuit symbol, (b) Truth table and (c) Timing diagram of a NAND gate.



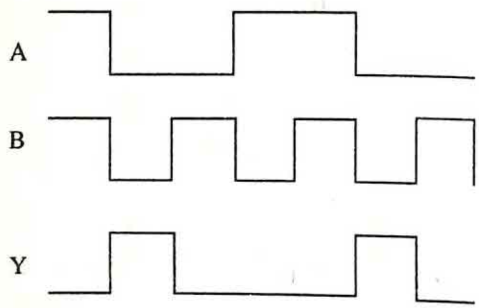
SYMBOL OF 2-INPUT NOR GATE

(a)

TRUTH TABLE OF 2 INPUT NOR GATE

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

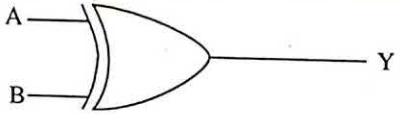
(b)



TIMING DIAGRAM

(c)

Figure 14.43: (a) Circuit symbol, (b) Truth table and (c) Timing diagram, of a NOR gate.



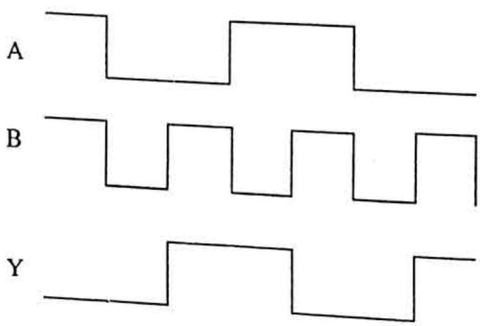
SYMBOL OF 2-INPUT X-OR GATE

(a)

TRUTH TABLE OF 2 INPUT X-OR GATE

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

(b)



TIMING DIAGRAM

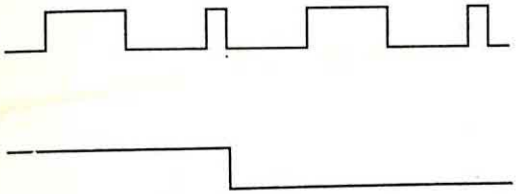
(c)

Figure 14.44: (a) Circuit symbol, (b) Truth table and (c) Timing diagram, of a X-OR gate.

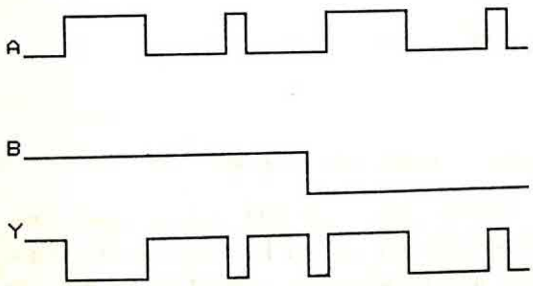
Example 14.12: Draw the output waveform of an XOR gate with the given inputs (Fig. (a)).

Answer: The characteristic of a two input XOR gate is, that the output is low for same inputs and high for different inputs. In Fig.(a) when the input at B is high, the out-

put is the complement of the A input and when, the B input is low, the output is the same as the A input (Fig.(b)).



(a)



(b)

14.6.3 Arithmetic circuits

The arithmetic operations performed directly by the computer are simply addition and subtraction. More complex operations such as multiplication, division, etc., are performed by repeated addition and subtraction. Circuits (adders) that perform these operations are described here. Thus, we can see how gates can be combined to achieve a bigger circuit with a useful function.

(i) Half adder

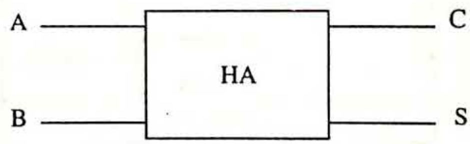
Binary addition can be summarised as shown in the table.

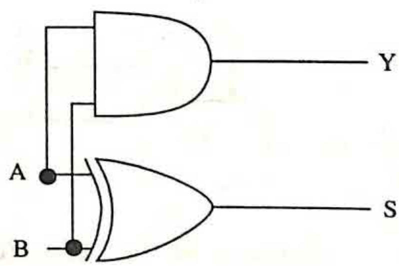
$0 + 0 = 0$

$0 + 1 = 1$

$1 + 0 = 1$

$1 + 1 = 0$ and carry 1





(a)

Half adder

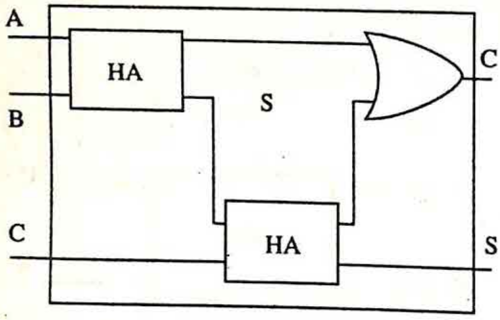
A	B	SUM	CARRY
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

(b)

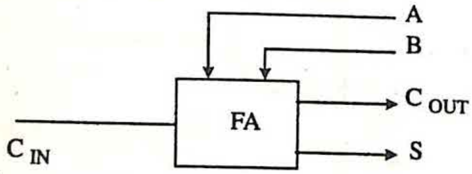
Figure 14.45: (a) Circuit and (b) Truth table for a half adder.

A half adder can add two bits and is shown in Fig. 14.45. Notice that it has two inputs and two outputs. A half adder can be realized using AND and X-OR gate as shown in the figure.

Table shows the Truth table. This circuit can add 2 bits. What we need now is a circuit that can add 3 bits at a time so that we



(a)



(b)

TRUTH TABLE

A	B	C	SUM	CARRY
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

(c)

Figure 14.46: (a) Circuit, (b) Symbol and (c) Truth table for a full adder.

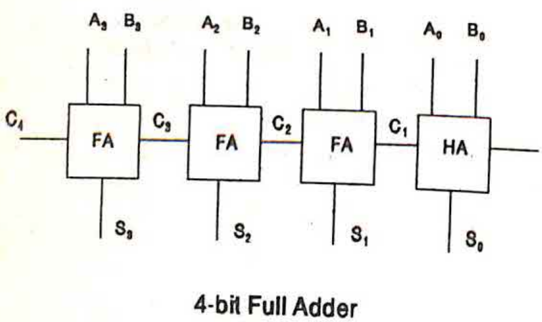


Figure 14.47: Circuit for a 4 bit full adder.

can add the carry also. This can be done by full adder.

(ii) Full Adder

A Full adder is shown in Fig. 14.46(a). Again there are two outputs SUM and CARRY but three inputs A,B,C. The table shows the Truth Table.

Let us take three bits and see how these three bits are added by this circuit. First two bits are given to A and B inputs and the third bit to C. The result will be as shown in the Fig. 14.46(b).

Full adder is the back bone of arithmetic circuits because it can be cascaded to add large binary numbers. Let us now design a 4-bit binary adder using the full and half adder circuits. Fig. 14.47 shows a 4-bit binary adder. A3, A2, A1, A0 is a 4-bit binary number A which is to be added to B3, B2, B1, B0, another 4-bit binary number B. The result will be S3, S2, S1, S0 and C4 (carry). Let us take an example and see how this works. Suppose A = 1101 and B = 1011. Then A+B is given by

$$\begin{array}{lllllll} A & = & A3 & A2 & A1 & A0 & 1101 \\ B & = & B3 & B2 & B1 & B0 & +1011 \\ \hline C4 & & S3 & S2 & S1 & S0 & 1\ 1000 \end{array}$$

Using column addition we get 11000. The

binary adder electronically comes up with the same answer. Here Half adder adds 1 (A0) and 1 (B0) to give a sum of 0 (S0) with a carry of 1 (C1). This carry goes to the carry input of the next Full adder, where 0 (A1) and 1 (B1) are the other two inputs. Addition of C1, B1 and A1 gives a sum of 0 (S1) and carry of 1 (C2). Now the next FA adds 1 (A2) 0 (B2) and 1·(C2) to give a sum of 0 (S2) with a carry 1 (C3). The last FA adds 1 (A3) 1 (B3) and carry 1 (C3) giving a sum of 1 (S3) with carry 1 (C4). Thus the result of adding the two 4 bit numbers A(1101) and B(1011) is 11000. Adders like other digital blocks, are available in IC form.

(iii) Subtraction

A digital circuit that will perform the logical operations required for binary subtraction can readily be designed. However, it is far more convenient to have a single circuit which can do both addition and subtraction. The most popular method of subtraction is through the addition of 2's complement. We will describe this method now. Before we describe 2's complement we need to know 1's complement of a binary number.

The 1's complement of a binary number is obtained by changing all 1s to 0s and 0s to 1s. In other words, by complementing each bit. For example, complement of 0000101 is 1111010. 2's complement is obtained by adding 1 to the 1's complement.

The complements of the binary numbers can be used to carry out subtraction. For instance let us subtract 101 from 111. This is illustrated in Table 14.3. Note that we have crossed out the final carry in the 2's complement method. What remains, is the answer in binary and is the same as that obtained by conventional subtraction. We can use the carry bit in our 2's complement method as the indicator for the sign of the answer. When the carry is 1, the answer is positive and is in binary. What will be the answer if we subtract 7 from 5 in this method?. The subtraction by the addition of 2's compliment will be 110 with no carry. When there is no carry, the answer is negative and is the 2's complement of the actual number. Hence it has to be converted to binary by taking the 2's complement again.

Table 14.3:

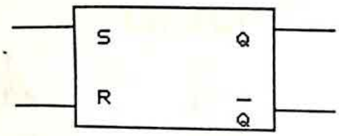
7	111	111
−5	−101	+011
2	010	010
Binary addition		Subtraction by the addition of 2's complement (neglect carry)

*14.6.4 Flip flop

We have so far learnt about the different types of gates and now, we will learn about *Flip flops* which are also important building blocks in digital electronics. Flip flops (FF) are used to make *counters* and *registers* which are used to count and store numbers, respectively. Memories, analog to digital converters and digital to analog converters are the other important digital blocks. Here we will learn elementary ideas about counters and registers. More about digital electronics can be studied in the higher classes.

(i) RS flip flop

The simplest flip flop (FF) is the RS flip flop. The circuit symbol is shown in Fig. 14.48(a). It has two inputs R and S and two outputs Q and $\overline{Q}$. If the S(*set*) input is high and R(*reset*) input is low, Q goes high and



(a)

RESET	SET	OUTPUTS	
R	S	Q	$\bar{Q}$
0	0	LAST VALUE★	
1	0	0	1
0	1	1	0
1	1	NOT ALLOWED	

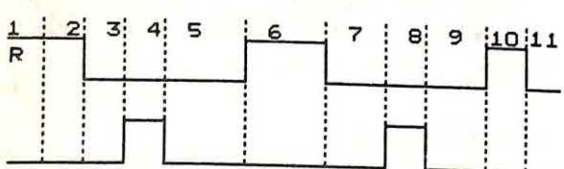
★ No input is applied and therefore the output retains its previous value.

(b)

Figure 14.48: (a) Circuit symbol and (b) Truth table for a RS flip flop.

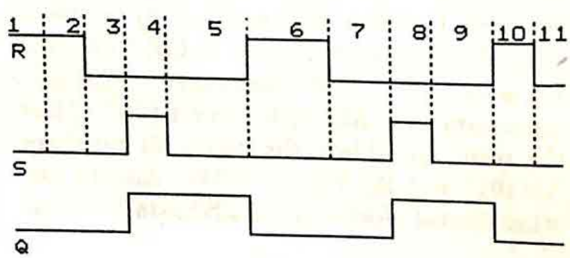
Q is low. Like wise, when the R input is high and S input is low then Q goes low and Q is high. The truth table (Fig. 14.48(b)) summarises the input output possibilities for RS flip flop.

Example 14.13: Draw the output waveform for the given inputs at R and S (Fig. (a)) using the truth table.



(a)

Answer: In the time slots 1 and 2, R is high and S is low. So, Q is reset to zero. In the time slot 3, R and S are zero, and the output remains in the same state as before - i.e.,at zero. In the time slot 4, $R = 0, S = 1$ and hence, Q is set to high. In the same fashion continuing the argument for all the eleven time slots, we get the output waveform as shown in Fig.(b).



(b)

When there is no input to R and S, then the output will remain the same. i.e., the previous value will be retained. When both R and S are high, the output is not defined. Hence this is not allowed.

A simple RS flip flop can be converted into a clocked RS flip flop as shown in Fig. 14.49. Often, in digital systems a pulse generator, called the clock, keeps all flip flops in the circuit in step with one another.

When the clock input is low, both the AND gates are disabled(i.e, output of AND gates is low). This ensures that $R = S = 0$,

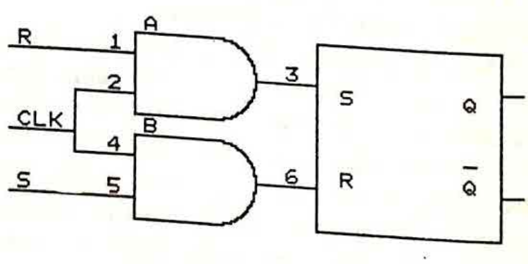
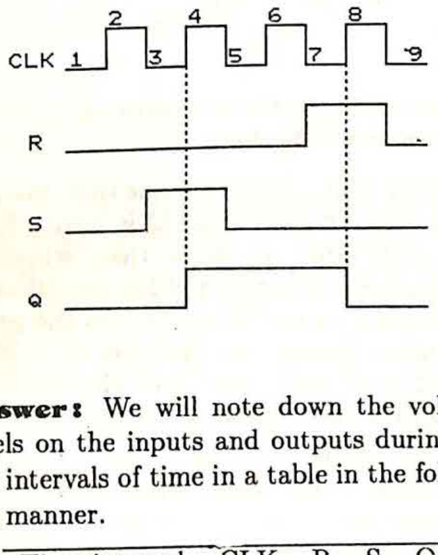


Figure 14.49: Circuit symbol for a clocked RS flip flop.

which means that Q and $\overline{Q}$ retain their previous values. When the clock input goes high, both AND gates are enabled, the S' and R' signals reach the RS inputs of the flip flop. Therefore, the clocked RS flip flop cannot change its output states until the clock signal arrives.

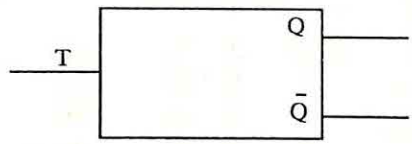
Example 14.14: Figure below represents the timing relationship between the clock, R, S inputs and the output Q. Derive the truth table using the timing relationship.



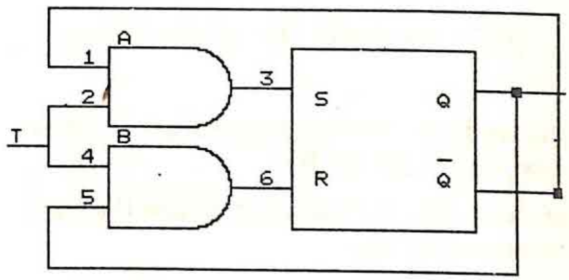
Answer: We will note down the voltage levels on the inputs and outputs during all the intervals of time in a table in the following manner.

Time interval	CLK	R	S	Q
1	0	0	0	0
2	1	0	0	0
3	0	0	1	0
4	1	0	1	1
5	0	0	0	1
6	1	0	0	1
7	0	1	0	1
8	1	1	0	0
9	0	0	0	0

From the above table we can clearly see that, when the Clock level is zero, any change in inputs on R and S will not have any effect on the output. Similarly, when both the inputs are zero, even when the clock is high, the output will continue to be the same.



(a)



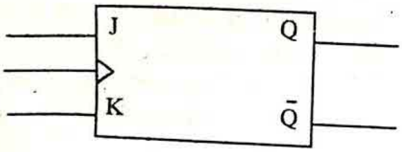
(b)

Figure 14.50: (a) Circuit symbol and (b) circuit diagram for a toggle (T) flip flop.

(ii) T flip flop

Fig. 14.50(a) shows one way to build a toggle flip flop. Its output states toggle (alternate) upon each input T pulse.

Suppose $Q = 1$ and $\overline{Q} = 0$. Then the lower AND gate is enabled. When the T pulse arrives, the flip flop is reset and the output becomes $Q = 0$ and $\overline{Q} = 1$. This enables the upper AND gate. When the next T pulse arives, the flip flop is set so that the output changes back to $Q = 1$ and $\overline{Q} = 0$. Thus each incoming T pulse alternately sets and resets the S and R inputs and the flip flop toggles . It can be seen from the figure that the output has half the frequency



(a)

TRUTH TABLE

CLK	J	K	Q
0	X	X	NC
1	0	0	NC
1	0	1	0
1	1	0	1
1	1	1	TOGGLE

NC = No change, X = Either 1 or 0.

(b)

Figure 14.51: (a) Circuit symbol and (b) circuit diagram for a JK flip flop.

of the input, i.e., T flip flop divides the input frequency by two.

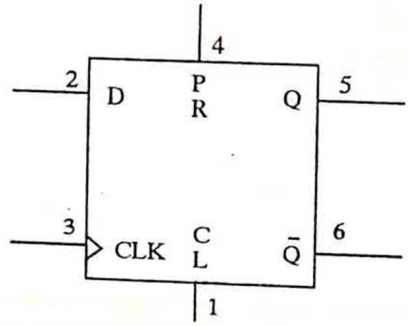
(iv) JK flip flop

A JK flip flop is a combination of a clocked RS flip flop and a T flip flop. Its circuit symbol and the truth table are given in Fig. 14.51.

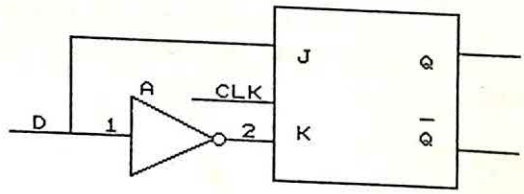
The JK flip flop can be used as a simple toggle flip flop (J=K=1) or the J and K inputs can be used to set or reset the output states as required.

Example 14.15: The waveforms representing the Clock, J and K inputs are given in Fig. (a). Draw the output waveform for the JK flip flop by taking the waveforms in Fig. (a) as inputs and using the truth table.

Answer: During the time slot 1, the clock = 1, J=0, K=0. So, the output retains its previous value (Let us assume that, the previous value of the output Q is zero). From the truth table, we know that, when the clock is low, the inputs will have no effect on the output. i.e the output retains the previous value. During the time interval 3, clk = 1; J=1; K=0 and Q is set to high (Check the output for the corresponding inputs given, in the truth table). During the time slots 4, 5, 6, there is no change in the output. During the time slot 7, clk=1, J=0 and K=1. This condition resets the output to zero. During the intervals 8, 9, and 10 the output remains the same. During the time intervals 11, 13, and 15, the output toggles as J=K=1. The output waveform is given in Fig. (b).



(a)



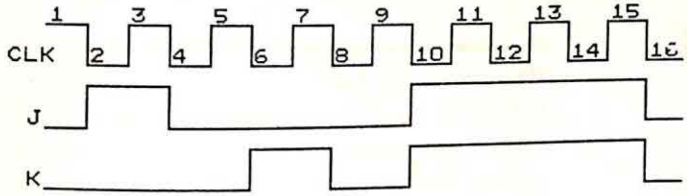
(b)

Figure 14.52: (a) Circuit symbol and (b) circuit diagram for a D flip flop.

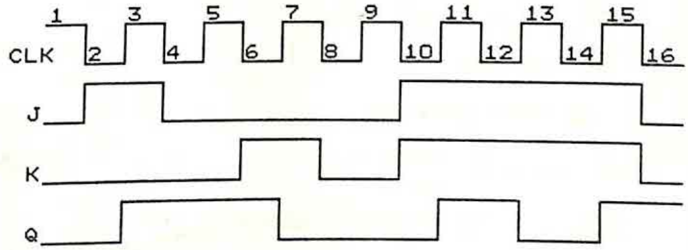
(v) D flip flop

The D flip flop is a modification of the JK flip flop. The circuit symbol for a D flip flop is shown in Fig. 14.52.

The D input is applied to the J input and.



(a)



(b)

Figures (a) and (b) for Example 14.15

its complement is internally applied to the K input.When the clock signal arrives, the data (i.e., input at D) appears at Q.

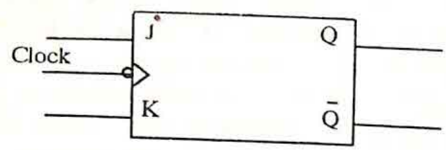
(vi) Edge-triggered flip flop

With this kind of flip flop the output changes state on the leading edge or trailing edge of the clock pulse as specified by the manufacturer. This is accomplished by modifying the input circuit with gates. It is enough to know,at this stage, that there are flip flops which trigger on the raising edge or falling edge of the input clock signal.

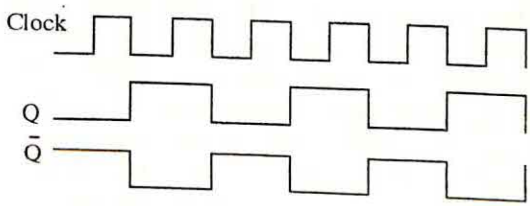
Example 14.16: Draw the time related waveform at the Q and outputs of the arrangement shown in Fig. (a), in response to the clock applied. Assume that initially Q state is at 0.

Answer: Figure shows the JK flip flop connected in toggle mode (J=K=1). Also as shown in the figure, the symbol -0> indicates that the output changes its state at every falling edge of the clock. The output waveform starting with $Q = 0$ and $\overline{Q} = 1$ is given in Fig. (b).

Observe that the period of the output waveforms at Q and $\overline{Q}$ are exactly two times that of the clock input applied. This circuit can be used as a frequency divider (The circuit divides the input frequency by two so that the frequency of the output waveform is half of the input waveform).



(a)



(b)

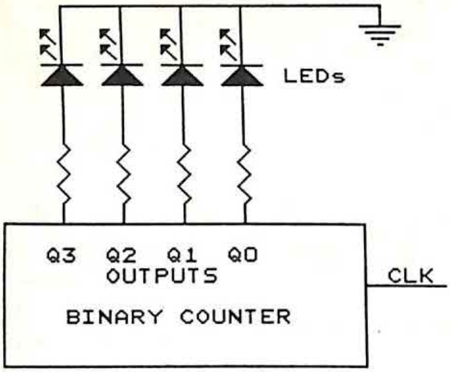


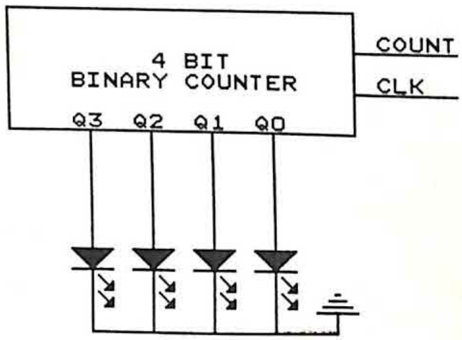
Figure 14.53: A binary counter.

(vii) Application of flip flops

(a) Counters:

A binary counter counts the number of input clock pulses as a binary number. In counters , JK flip flops are used in the toggle mode i.e., J=K=1 always. The outputs of the flip flops change states with each clock pulse. The counter shown in Fig. 14.53 is called a 4-bit counter. It has a clock input, a control input and the binary outputs. It counts the number of clock pulses for the duration of the control (count) pulse. With each clock pulse the output increases by one i.e.,Initially, all outputs will be zero. After the first clock the output reads 0001. After the second clock pulse the pulse output reads 0010 and so on till it reaches 1111. When the output is 1, the LED glows. This counter is called an up-counter because the counter counts up to the next higher number for each input clock pulse. A counter can also be made which counts down from a given number to zero.This is called a down-counter. Counters that can count both up and down are called up- down counters. All these different kinds of counters are available as IC's and can be readily obtained and used.

Example 14.17: **What is the number displayed by the LEDs when the pulse applied on the count pin is $100\mu s$ and the clock period is $10\mu s$, in the figure below. (Assume that the counter is initially showing zero)**

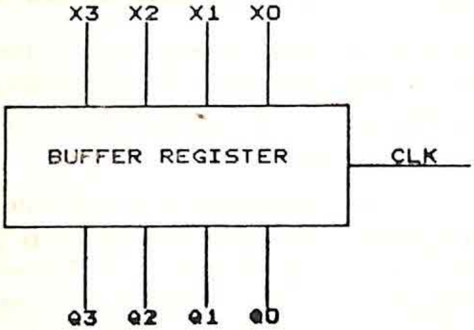


Answer: When the count pin is enabled (i.e, the logic level on the count pin is 1) the counter starts accepting the clock pulses and increments the count by one for every clock pulse. When the logic level on the count pin is zero, the counter is disabled and the counting is stopped.

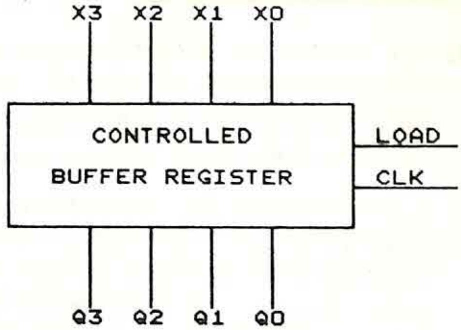
Assuming that the counter is initially set to zero, when the count is enabled by a pulse of $100\mu s$ duration, $100/10 = 10$ clock pulses will be accepted by the counter. So, the display shows 1010 (binary) at the end of $100\mu s$.

(b) Registers:

A register is a group of D-flip flops which can be connected in series to store a binary number, one flip flop for each bit of the binary number. Again, registers are available as IC's. A buffer (or, storage) register is the simplest kind of register- all it does is to store a binary number. Fig. 14.54(a) shows a 4 bit buffer register When the clock signal arrives, the four X bits $X_3 - X_0$ are loaded into register and the output reads



(a)



(b)

Figure 14.54: (a) A 4-bit buffer register and (b) A 4-bit buffer register with 'load' pin.

$Q_3 = X_3, Q_2 = X_2, Q_1 = X_1, Q_0 = X_0$ (i.e., $Q = X$). In real IC registers, a 'Load' pin is provided in addition to the Q, X, and Clk pins as shown in Fig. 14.54(b).

We can give the 4 input bits at X_3, X_2, X_1, X_0. When load pulse is given, X (= X_3, X_2, X_1, X_0) is transferred to the output Q (= Q_3, Q_2, Q_1, Q_0) on the arrival of the clock.

Summary

1. The structure of a crystal is specified by its unit cell which repeats itself in three different directions in space to form the entire crystal. The unit cell is a parallelopiped characterized by three lengths (lattice constants) and the angles between them. Simplest structures are cubic (e.g. NaCl, Cu) and hexagonal (e.g. Zn, ZnO).

2. A free atom has well-defined energy levels; e.g. Si has the electronic configuration: $1s^2 2s^2 2p^6 3s^2 3p^2$. These energy levels, especially those of the outer shell electrons, get modified when atoms are brought close, as in a crystal. For an assembly of N free Si atoms, there are $2N$ electrons completely filling the outer S-level and $2N$ electrons in the outer P-level which can accommodate $6N$ electrons. In the Si crystal, however, the $8N$ levels split into two groups, each of very closely spaced energy levels (bands): the completely filled valence band containing $4N$ levels, and the unfilled conduction band also containing $4N$ levels. The two bands are separated by an energy gap.

3. In metals, the conduction band is partially filled. In insulators, it is empty while the valence band is completely filled, and the two bands are separated by a wide energy gap; semiconductors have a similar band structure but the energy gap is considerably smaller. These features can be used to explain why metals are good conductors of electricity, insulators are poor conductors and the resistance of semiconductors decreases with increasing temperature.

4. When an electron is removed from the valence band, the vacancy so created is called a *hole*. A hole has a positive charge e and the same mass as the electron; its energy is higher the farther below it is from the top of the valence band. In pure or intrinsic semiconductors, the electron density in the conduction band n_e and the hole density in the valence band n_h are equal. A semiconductor with (desirable) impurity atoms is called a doped or extrinsic semiconductor.

5. Silicon doped with pentavalent atoms (e.g. As, Sb, P) is a n-type semiconductor. Here four of the five valence electrons are shared in covalent bonding with Si, while the fifth is "donated" to the host crystal. The donor energy level lies a little below the bottom of the conduction band. Therefore, at any non-zero temperature, a sizable fraction of the donor electrons is in the conduction band. Usually for this case, $n_e >> n_h$.

   Silicon doped with trivalent atoms (e.g., In, B, Al) is a p-type semiconductor. Here for completing the In-Si bond, a hole is created in the covalent bonds in the crystal. The acceptor energy levels lies a little above the top of the valence band. For this case, $m_h >> n_e$.

   These features enable us to modify the electrical properties of semiconductors dramatically by doping them with small amounts of suitable materials. In a semiconductor, the current flow is due to both, free electrons in the conduction band and holes in the valence band. The resistivity of a semiconductor depends on the electron and hole densities and their mobilities.

6. In a pn junction, diffusion of electrons from n region to p region, and of holes from p to n result in an electric field (or potential barrier) at the junction which greatly checks further diffusion. The vicinity of the junction has no free charge carriers (depletion region). In equilibrium, the small diffusion current is balanced by the drift current. The latter arises due to minority carriers (electrons from p to n; holes from n to p) for which the direction of the barrier field is favourable.

   In a forward biased pn junction, the external electric field opposes the barrier field. This causes the diffusion current to increase; when the forward bias voltage equals or exceeds the barrier potential, the resistance is small. The drift current is not much affected by the bias voltage. The depletion region is reduced.

   In a reverse biased pn junction, the diffusion current decreases and the depeletion region becomes larger. The drift current remains small and unchanged at first. After a certain reverse biase voltage, the drifting electrons have large eneough knietic energy to break covalent bonds and produce an avalanche of electrons. (reverse breakdown).

   The unidirectional characteristic of a pn junction (conducting when forward biased and non-conducting when reverse biased) is used in converting AC into DC (rectification). In a half-wave rectifier, only the positive half cycle of the input appears in the output; in a full-wave rectifier (which employs two pn junction diodes suitably) output appears for both the half cycles of the input.

7. The juction transistor consists of three doped semiconductor layers in the pnp or npn configuration (Emitter - Base - Collector). Consider a npn transistor with forward-biased B-E junction and reverse-biased B-C junction. The sandwiched region (Base) is thin and lightly doped, and the B-C junction has a greater area than the B-E junction. The base current $I_B(\sim \mu A)$ is, therefore, much less than the collector current $I_c(\sim mA)$ which means the emitter current

   $$I_E = I_B + I_C \simeq I_C.$$

   Since the B-E junction is forward-biased, its resistance R_{BE} is much less than that of the reverse-biased B-C junction. Since the same current ($\sim I_C$) flows through both, there is power gain. The gain can be made to appear across a load of large enough resistance R_L that does not affect I_C greatly. ($I_C^2 R_{BE} << I_C^2 R_L$)

8. In digital circuits only two values (represented by 0,1) of the input and output voltage are permissible. A logic gate is a digital circuit with certain logical relation between the input and output voltages. In the OR gate, the output is 1 if any of the input is 1. In the AND gate, the output is 1 only when both the inputs are 1. In the NOT gate, the output is 1 if the input is 0 and vice-versa. The first two gates can be realized using pn junction diodes; the NOT gate requires a transitor. Various combinations of these three basic logic gates lead to complicated digital circuits for different applications.

Exercises

14.1 Draw the output waveform for the input sinewave given in Example 14.2, when the diode's direction in the half wave rectifier is reversed (see Fig.14.22(a))

14.2 What is the input voltage needed to maintain 15V across the load resistance R_L of 2K, assuming that the series resistance R is 200 Ω and the zener requires a minimum current of 10mA to work satisfactorily. (Use the figure of Example 14.3). What is the zener rating required.

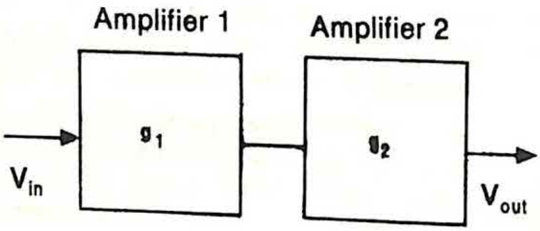
14.3 For a transistor connected in common emitter mode, the voltage drop across the collector is 2V and β is 50. Find the base current, if R_c is 2K.

14.4 In a transistor connected in common emitter mode, $R_c = 4\text{K}$, $R_i = 1\text{K}$, $I_c = 1\text{mA}$, and $I_b = 20\mu\text{A}$. Find the voltage gain.

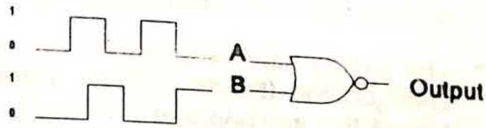
14.5 Draw the circuit to (a) forward biased diode. (The supply is 3V, 100mA battery). If the diode is made of silicon and the knee voltage is 0.7V, and a current of $20mA$ passes through the diode, find the wattage of the resistor and the diode.

14.6 In half wave rectification if the input is 50Hz, what is the output frequency. What is the output frequency of a full wave rectifier for the same frequency.

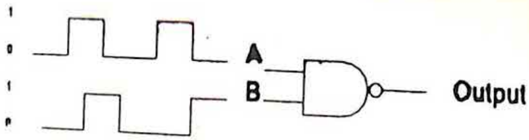
14.7 Two amplifiers are connected as shown in the figure. Find V_0 if $g_1 = 20g_2 = 10$ and $V_{in} = 50$ mV. $V_z = 10$V for each diode.



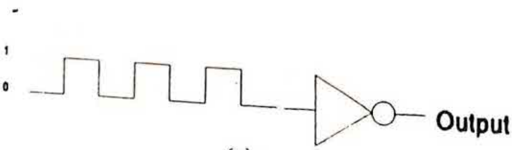
14.8 Write the output waveform for the gates shown in figures (a), (b) and (c).



(a)



(b)



(c)

14.9 Perform binary addition on the following sets of numbers.

(a) 110010 and 111101

(b) 101010 and 010101

(c) 111111 and 000001.

14.10 Complement (1's) the following binary numbers.

(a)	110010	(c)	101010
(b)	111101	(d)	111111

14.11 Find the 2's complement of the following numbers.

(a)	111111	(c)	101010
(b)	000000	(d)	001001

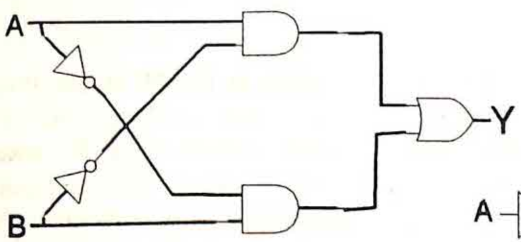
14.12 Perform subtraction on the following sets of numbers.

(a) 110010 and 111101

(b) 010101 and 101010

(c) 1000001 and 1111111

14.13 Write the truth table for the circuit given in the figure.



Additional Exercises

14.14 The internal voltage reference of IC3085 voltage regulator is 1.5V. The value of R used in the potential divider circuit is 80K. By using a potentiometer (variable resistance) in the place of R_2, instead of a fixed resistor, we can vary the output by changing the R_2 value. Find the value of the potentiometer to be used in order to have a variation of 10 to 20V at the output.

14.15 The input and output voltages of a transistor amplifier at different frequencies are shown in the table.

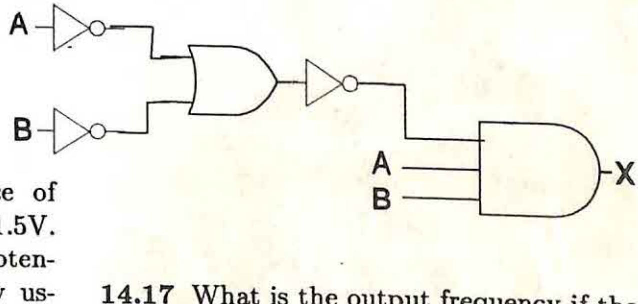
Frequency (Hz)	V_i (V)	V_0 (V)	Frequency (Hz)	V_i (V)	V_0 (V)
500	1	2	2500	1	4
1000	1	3	3000	1	4
1500	1	4	3500	1	3
2000	1	4	4000	1	2

(a) Find the gain of the amplifier at all the frequencies given.

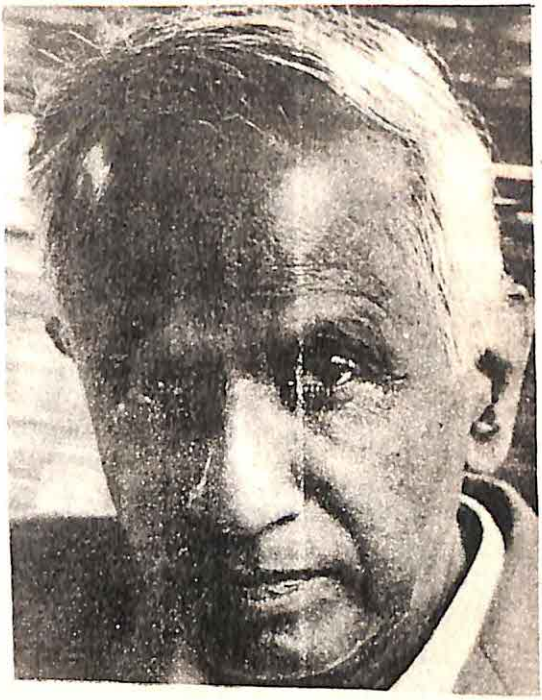
(b) Draw a graph for frequency Vs gain on a graph sheet.

(c) Find the cut off frequencies $f_1$and f_2.

14.16 Find the output X of the following circuit if the inputs are: $A = 0, B = 0; A = 0, B = 1; A = 1, B = O; A = 1, B = 1;$



14.17 What is the output frequency if the clock input to a JK flip flop, both the inputs of it connected to high ($J = K = 1$), is 200KHz.



S. Chandrasekhar (1910-1995) Born in Lahore, India 1910. After his education in India and Cambridge University, England, he was with the University of Chicago, USA since 1937. He made extensive and important contributions in the areas of stellar strcutre and evolution, atmospheres of stars, stellar dynamics and the theory of black holes. On the basis of his theory of white dwarfs, he derived the well-known Chandrasekhar limit for their masses. He was awarded the Nobel Prize in 1983.

Nicolaus Copernicus (1473-1543) Born in Poland, Copernicus studied law and medicine inaddition to astronomy. For most of his life, he was a canon at Frauenburg Cathedrals. He revolutionised astronomy by his heliocentric theory of the solar system which he described in his book 'De revolutionibus orbium celestium'. On the basis of his theory, he demonstrated how to calculate accurately planetary position and distances from the sun. He also explained phenomena such as the occurrence of seasons and the apparent retrograde motion of planets. He was aware of the immense distances of the stars. The Copernican theory paved the way for the great discoveries by others like Kepler and Newton.

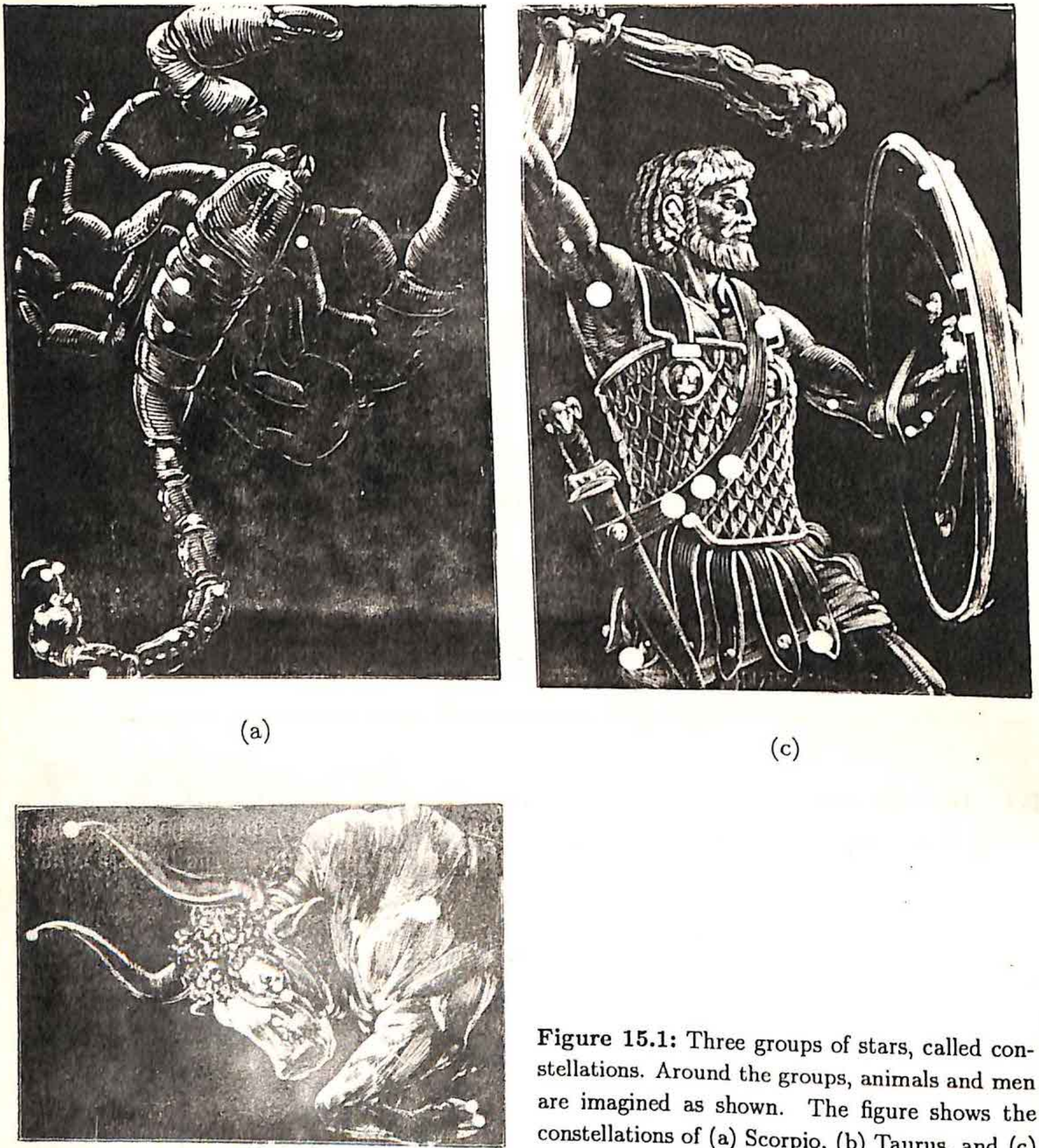
CHAPTER **15**

The Universe

15.1 Introduction

The night-sky with its splendour and mystery has been a source of perennial fascination to every sensitive person. Artistis, poets and philosophers have drawn inspiration from it. And scientists have constantly tried to understand the nature and behaviour of celestial objects visible in the sky. As a result of centuries of such investigation, we possess today a sufficiently good understanding of the universe at large as revealed by the heavens around us. But, the picture of the universe conceived by man has been a changing one. This picture has been the result of close interaction between observation and theory, between facts discovered and models created to fit those facts together. The accuracy of astronomical observations has improved with time especially with the development of telescopes and other instruments. Similarly, our knowledge of the laws of nature has progressively increased. As a result our concept of the universe has evolved through the ages.

On a clear night, if unhindered by city light, we can see several thousand stars. If we watch the stars for a period of time, we can see that they rise in the east, follow circular trajectories and set in the west, just as the sun does during the day. The obvious explanation would be that the stars actually went round the earch or the *celestial sphere*, in which the stars seem to be imbedded, turned round the earth. It required the keen intellect of astronomers like Aryabhata (fifth century AD) to realise that it was the



(a) (c)

(b)

Figure 15.1: Three groups of stars, called constellations. Around the groups, animals and men are imagined as shown. The figure shows the constellations of (a) Scorpio, (b) Taurus, and (c) Orion.

earth's rotation about an axis that caused the apparent circular motion of the stars. A verse from his book *Aryabhatiya* reads: 'Just as a man moving in a boat downstream sees the trees on the river bank go in the opposite direction, so it is with stars which ap-

pear to move in the westerly direction when in fact they are fixed in space'.

If we observe, while in motion, two objects at different distances from us they seem to move relative to each other. This is due to the phenomenon of 'parallax'. The farther the two objects from us, the less this relative apparent motion between them. The stars are so far away from us that the earth's rotation does not give rise to this apparent motion among them. Their relative positions in the sky are thus fixed and therefore they are called the 'fixed stars'. The ancient Indian astronomers named some of the stars after sages and mythological figures. Many of the names of the stars commonly used today, e.g. Aldebaran, Algol etc., were originally given to them by the Arabs. Similarly, the Greeks could imagine in the fixed patterns of stars the images of their epic heroes and heroines, beasts and monsters. 'Constellations' or groups of stars still bear the names given to them by the Greeks: Hercules, Andromeda, Hydra and so on. Stars in a given constellation may be at different distances from us. We cannot readily ascertain this without careful measurements, because the stars are so far away. The constellations provide convenient reference points to locate celestial objects and astronomical events.

Against the backdrop of the fixed stars, a few celestial objects seemed to move in more complicated ways when observed over the period of a year. These were the planets, a term which meant 'wanderers' in Greek. In order to explain the motions of planets, the Greek astronomers assumed that each planet went round in a small circle, known as the epicycle, the centre of which itself followed a circular path around the fixed earth. As observations became more accurate, the models explaining them became more complicated with increasing number of circles added at each stage. This *geocentric model*,

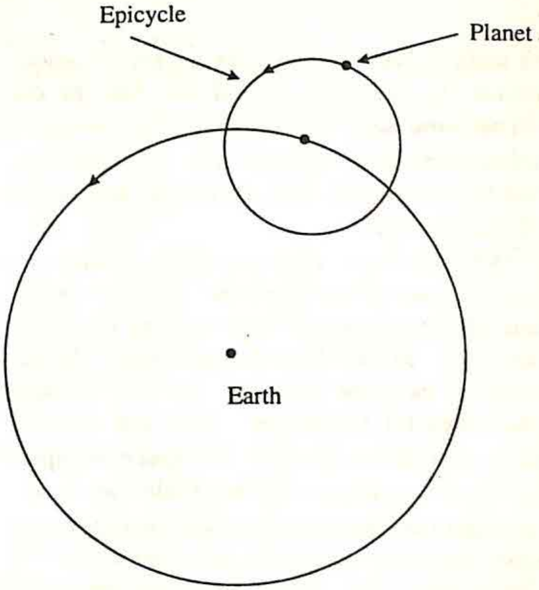


Figure 15.2: The geocentric model of planetary motion.

in which the earth stood still and all the celestial bodies revolved around it, was put forward in its most refined version by the Greek astronomer Ptolemy in the first century AD.

The geocentric theory was widely accepted for fifteen centuries, until Copernicus (1473-1543) proposed his *heliocentric theory* in which the sun was at rest and the planets including the earth went around it. This was an extraordinarily bold step conceptually and simplified considerably the model of the solar system. Moreover, it opened up the path that was followed by great discoverers like Galileo (1564-1642), Kepler (1571-1630), Newton (1642-1727) and many others.

We have already studied the three laws of planetary motion of Kepler that described how the planets moved around the sun. Newton's law of gravitation explained how sun's gravitational pull was responsible for Kepler's laws. This also introduced the notion of universal laws that were applicable to all bodies in nature. It is with the help

of such universal laws that we try to understand the universe around us. On the observational side, the steady development of telescopes, starting from that of Galileo, has led to continuous incrase in our knowledge of the heavens.

We now know that the solar system is a minute part of the universe. Our sun is but one of the billions of stars forming the milky way. The Mulky Way is one among the billions of galaxies that can be seen through the powerful telescopes. And the universe may extend far beyond the space occupied by these galaxies. As we shall see, in describing the universe, we shall encounter not only immense distances but also extraordinarily long time scales involving the evolution of the universe.

In the study of the universe, scientists jave to contend with several restrictions. For example, observations may be difficult to make because of the immense disances of the heavenly bodies from the earth. One can only observe astronomical phenomena but cannot perform experiments at will. Further, what we observe may be only part of what is actually happening. In spite of all this, it is an amazing fact that so much is known about the universe. Moreover, in understanding what we observe we use universal laws deduced from terrestrial experiments. Predictions based on these laws can be verified even in the case of astronomical phenomena. This shows that the laws of nature are valid over vast distances and long stretches of time, thereby making them truly universal. The universe is nature's vast laboratory in which all possible physical phenomena are at work producing the spectacular effects we observe.

In the following sections, we shall climb the universal ladder step by step. We shall try to glimpse the solar system, the Milky Way, the galaxies and finally the large scale structure and evolution of the universe as a whole.

15.2 The solar system

"The sun sits as upon a royal throne, ruling his children, the planets, which circle round him", wrote Copernicus. Our solar system is dominated by the sun, all other objects like planets, comets and asteroids are bound to it by gravitation. Nearly 99.9 per cent of the matter in the solar system is accounted for by the sun itself, the planets making up most of the rest. Let us examine briefly the constituents of the system.

The Sun

The sun is a typical example of a star. It is a hot, radiating sphere of gas - predominantly hydrogen. We shall return to the energy production within the sun later. Here we note that the temperature of the outer visible region is about 6000 K. The visible part has a diameter of about 1,400,000 km, but sun's invisible thin gaseous envelope extends far beyond this. The sun's mass is about 2×10^{30} kg which is more than 300,000 times that of the earth. Energy reaching all parts of the solar system in the form of light and heat is derived from the sun.

The planets

There are nine planets revolving around the sun. Six of them, namely the Earth, Mercury, Venus, Mars, Jupitor and Saturn, have been known from ancient times. The other three - Uranus, Neptune and Pluto - were discovered later after the advent of the telescope. Compared to the sun, the planets are quite small and relatively cool. They are not self-luminous and shine only by reflecting the sunlight. We may note that the existence of the planet Neptune had been predicted even before it was discovered by observation on

the basis of the gravitational disturbance it produces in the orbit of Uranus.

Distance and size: The measurement of distances to the planets from the earth was discussed at the beginning of class XI (Physics, Vol I, Part I, Chapter 2: Physical World and Measurements). This is done by the *parallax method*. The planet is observed from two different observatories A and B on the earth (distance AB $= b$) at the same time. If θ is the angle betwen the two directons along which the planet is viewed at these two points, then the distance D from the earth to the planet is obtained from the relation

$$\theta = b/D \tag{15.1}$$

Notice also that θ is equivalently the angle subtended at the planet by the line AB. Here the assumption is that while D is very large, b and θ are small.

Similarly if d is the diameter of the planet and the angular size of the planet, i.e., the angle subtended by d at the earth is α then we have

$$\alpha = d/D. \tag{15.2}$$

The angle α can be measured from the same location on the earth (the angle between the two directions of the telescope when two diametrically opposite points of the planet are viewed). Since D is known, the diameter of the planet d can be determined from Eq. (15.2). These methods can be applied to measure the distance to the sun and its angular diameter as well.

Example 15.1: The moon is observed from two diametrically opposite points A and B on earth. The angle θ subtended at the moon by the two directions of observation is $1°54'$. Given the diameter of the earth to be about 1.2756×10^7 m, compute the distance of the moon from the earth.

Answer: We have $\theta = 1°54' = 114'$

$$= (114 \times 60)'' \times (4.85 \times 10^{-6})$$
$$= 3.3174 \times 10^{-2} \text{ rad},$$

since $1'' = 4.85 \times 10^{-6}$ rad.

Also $b = AB = 1.2756 \times 10^7$ m.

Hence from Eq. (15.1), we have the earth-moon distance,

$$D = \frac{b}{\theta} = \frac{1.2756 \times 10^7}{3.3174 \times 10^{-2}}$$
$$= 3.8452 \times 10^8 \text{m}.$$

Example 15.2: The Sun's angular diameter α is measured to be $1920''$. The distance D of the sun from the earth is 1.496×10^{11} m. What is the diamter of the sun?

Answer: Sun's angular diameter $\alpha =$ $1920''$

$$= 1920 \times 4.85 \times 10^{-6} \text{ rad}$$
$$= 9.312 \times 10^{-3} \text{ rad}.$$

Sun's diameter

$$d = \alpha D$$
$$= (9.312 \times 10^{-3}) \times (1.496) \times 10^{11}$$
$$= 13.93 \times 10^8 \text{m}.$$

The above method of parallax gives the distance of a planet from the earth. The relative distances of the planets from the sun, on the other hand, were determined way back by Copernicus himself .assuming the planetary orbits to be circles. His method is quite simple in the case of the *inferior planets*. This term describes planets that are closer to the sun than the earth, namely Venus and Mercury. The angle formed at

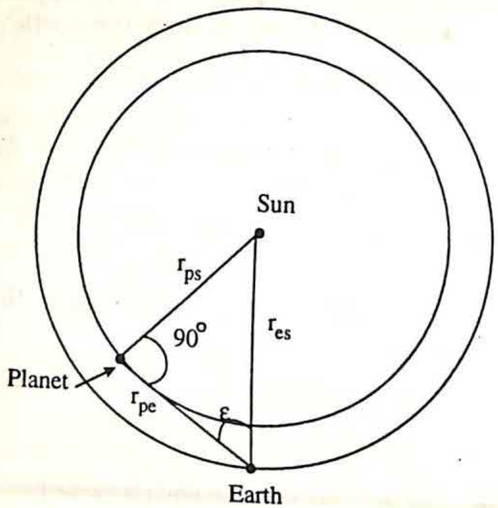


Figure 15.3: Determination of the distance of an inferior planet from the sun and from the earth.

the earth between the earth-planet direction and the earth-sun direction is called the planet's *elongation*. This is the angular distance of the planet from the sun as observed from the earth. When the elongation attains its maximum value ϵ, as in the figure, the planet appears farthest from the sun. It is easy to see that in this position the angle subtended by the sun and the earth at the planet is 90°. Therefore, it follows that the distance of the planet from the sun $r_{ps} = r_{es} \sin\epsilon = \sin\epsilon$ AU, where r_{es} is the average earth-sun distance which is called the *astronomical unit* (AU). We have thus determined the planet-sun distance in terms of AU. This distance can be expressed in kilometers, if AU is known in kilometers. To do this, we note that the earth-planet distance is given by $D \equiv r_{pe} = \cos\epsilon$ AU. Now if r_{pe} is determined directly in kilometers, then the value of AU in kilometers can be found from the above equation. In recent years, the radar has been made use of in the accurate measurement of the distance of nearby celestial objects such as Venus. Radio signals are bounced off the surface of Venus and these reflected pulses detected on the earth. It takes the signal, travelling with the speed of light c, a time $t = r_{pe}/c$ to cover the distance between the earth and Venus. Therefore, the time elapsed between the transmission of the radar signal and the reception of its echo is $2t$ which can be measured accurately. Since c is given in km/s., $r_{pr} = ct$ is found in kilometres. In turn the value of the astronomical unit is computed as $1\text{AU} = (r_{pe}/\cos\epsilon)$ km. It is thus found that 1 AU $= 1.496 \times 10^{11}$ m.

Example 15.3: In the case of Venus, the angle of maximum elongation ϵ is found to be approximately 47°. Determine the distance between Venus and the sun r_{vs} and the distance between Venus and the earth r_{ve}.

Answer: As has been shown above, we have respectively,

$$r_{vs} = \sin\epsilon \text{ AU} \;= \sin 47^\circ \text{ AU} = 0.73\text{AU}$$

and

$$r_{ve} = \cos\epsilon \text{ AU} \;= \cos 47^\circ \text{ AU} = 0.68\text{AU}.$$

Since 1 AU $= 1.496 \times 10^{11}$m, these distances can be expressed in meters also as $r_{vs} = 1.092 \times 10^{11}$m and $r_{ve} = 1.017 \times 10^{11}$ m.

Copernicus devised method to determine the distance from the sun in the case of *superior planets* also, i.e. planets whose orbits are larger than that of the earth. As known at the time of Copernicus, these are Mars, Jupiter and Saturn. The method is straightforward, but slightly more complicated than that used for the inferior planets.

Knowing the distance of any one of the planets from the sun, that of any other can be calculated. This is a consequence of Kepler's third law of planetary motion which states that the square of the period (T) of revolution of a planet round the sun is proportional to the cube of the semimajor axis (a) of the orbit. For two planets P_1 and P_2, therefore, we have

$$\frac{a_2^3}{a_1^3} = \frac{T_2^2}{T_1^2}. \qquad (15.3)$$

Periods can be ascertained by direct observation. Therefore, if a_1 is measured a_2 can be calculated.

Example 15.4: Starting from the results of Example 15.3, find the orbital period of Venus in days.

Answer: In Example 15.3, we determined $r_{vs} = 0.73$ AU. Taking this to be the average distance between Venus and the sun, we have $a_v = 0.73$ AU. Eq.(15.3) can be written as

$$\frac{a_v^3}{a_e^3} = \frac{T_v^2}{T_e^2}.$$

Substituting the value of a_v and setting $a_e =$ 1AU, $T_e = 1$ year, we get

$$T_v^2 = \left(\frac{a_v}{a_e}\right)^3 .T_e^2 = (0.73)^3 \text{ y}^2 = 0.39 \text{ y}^2.$$

Or, $T_v = 0.62$ y $= 226$ days.

The mean distances of the planets from the sun range from 0.39 AU (58 million km) for Mercury to 39.46 AU (5900 million km) for Pluto. Their periods of revolution round the sun range from 88 days for Mercury to 248 years for Pluto. The orbits of the planets are all approximately in the same plane. The maximum inclination occurs for Pluto whose orbital plane makes an ange of 17° with that of the earth. In diameter, the planets range from 3000 km (Pluto) to 1,43,000 km (Jupitor).

Mass: When we studied gravitation in class XI (Chapter 8) we considered Kepler's third law. If a mass m follows a circular orbit of radius R with period T around a central mass M, then

$$\frac{GMm}{R^2} = \frac{mv^2}{R} \text{ and } v = \frac{2\pi R}{T}.$$

It thus follows,

$$M = \frac{4\pi^2 R^3}{GT^2}. \qquad (15.4)$$

If R and T can be measured, then M can be calculated from Eq. (15.4). If the formula is applied to planetary motion around the sun, the mass of the sun can be determined. If the orbit is elliptical, R is replaced by the semimajor axis a. If the planet has a satellite, then the mass of the planet itself can be determined by the above method. All planets, except Mecruty and Venus, have satellites. In the absence of a satellite, a planet's mass has to be found out from its gravitational effect on another object passing by such as a spacecraft or a comet. Knowing the mass and the radius of a planet, one can immediately find its density. The masses of the planets, in terms of the mass of the earth, range from 0.002 (Pluto) to 318 (Jupitor).

Example 15.5: Assuming that earth's orbit around the sun is a circle of radius $R = 1.496 \times 10^{11}$m, compute the mass of the sun ($G = 6.668 \times 10^{-11}$ N m^2/kg^2).

Answer: Applying Eq.(15.4) to the sun-earth system with R given as above and $T =$

1 year = 3.156×10^7s, we obtain,

$$M_{sun} = \frac{4\pi^2(1.496 \times 10^{11})^3}{(6.668 \times 10^{-11})(3.156 \times 10^7)^2}$$
$$= 1.989 \times 10^{30}\text{kg}.$$

Atmosphere: In studying graviation, we have already seen how the nature of the atmosphere - its thickness and the molecules found in it - depends on the graviational pull of the planet and the average speed of a given molecule. These two factors depend respectively on the mass of the planet and the temperature of the atmosphere. The chemical composition of the planetary atmosphere can be ascertained from the spectrographic analysis of sunlight reflected from the planet. The molecules of the gases in the atmosphere absorb certain wavelengths producing dark lines in the spectrum. From the analysis of these lines the molecules present can be identified. From such studies it has been found that the main constituent of the atmospheres of Venus and Mars is carbon dioxide. Jupiter and Saturn have atmospheres that contain hydrogen, helium, methane and ammonia; those of Uranus and Neptune contain hydrogen, helium and methane. Becaue of its proximity to the sun, Mercury is too hot to possess an atmosphere. The composition of Pluto's atmosphere is not properly known.

We note that the planets are classifeid into two groups according to their physical properties: the terrestrial planets - Mercury, Venus, Earth and Mars (Pluto is also sometimes included) and the Jovian planets (named after Jupitor) - Jupiter, Saturn, Uranus and Neptune. The terrestrial planets are relatively small, dense and are probably made of rocks and metallic material. The Jovian planets are comparatively large, low in density and are predominantly composed of hydrogen and helium.

Other constituents

Comets: Comets are composed of frozen gases with nuclei of solid particles. They move around the sun in highly elongated elliptical orbits, spending most of the time far away from the sun. When a comet enters the solar neighbourhood, it gets heated and the vapourised gases and the particles form the head. A long trailing tail turned away from the sun is created by the radiation presure and the solar wind which consists of gas streaming out of the sun's atmosphere. While the heads of comets are 10,000 - 20,000 km in size, the tails can stretch to millions of kilometers. The mass of a comet is exceedingly small, about a billionth that of the earth. Each year five to ten new comets are discovered. Thus more than a thousand comets have been so far observed. There is possbily a large cloud of comets, numbering hundreds of billions surrounding the sun. Most of them have orbits beyond the outermost planets, so that they are never seen from the earth.

Asteroids: Asteroids, or minor planets, are minute planets most of whom are only a few kilometers across. Ceres, the largest of the some tens of thousands observable ones, is around 1000 km in size. Like the regular planets, the asteroids follow elliptical orbits nearly in the same plane as that of the earth. Most of them are confined to the space between the orbits of Mars and Jupiter. Some of them, however, come quite close to the earth.

Meteoroids: Minute objects orbiting the sun, meteroids are too small to be observed even with telescopes. They become visible when they enter earth's atmosphere and heat up due to friction as they plunge towards the earth. These 'shooting stars' are then known as 'meteors'. More than a hundred million

meteoroids are estimated to enter earth's atmosphere within a period of 24 h. Some occasionally survive their journey through earth's atmosphere and land on its surface. They are then called 'meteorites'.

15.3 The stars

The night sky is sprinkled with thousands of stars that appear like bright points of light. Astronomers, with their telescopes and other intruments, can determine a few characteristics of the stars like distance, brighness and the features in the specturm. Surprisingly, even with such meagre information one can learn a great deal about the stars - mass, size, temperature, chemical composition and their evolution from their birth to their inevitable death. We shall briefly outline these aspects of stars in general. but, first let us take a look at the star nearest to us, namely the sun.

The Sun, our nearest star: We are dazzled by the brilliance of our sun because of its relative nearness to us. In reality, it is only an average star in its size, mass and brightness. As mentioned earlier, the mean distance between the earth and the sun, the astronomical unit (AU), is equal to 1.496×10^8 km. This can be measured by the parallax method. When dealing with the large distances to the stars and other celestial objects it is conveneint to use the light travel time in defining the unit of distance measure. A light year (ly) is the distance travelled by light in one year so that ly = 9.46×10^{12} km = 63,240 AU. It takes about 8 min. for light to travel from the sun to the earth. Therefore our sun is 8 light minutes away from us.

We have seen how the mass of the sun can be measured by using Kepler's law from a knowledge of a and T of an orbiting planet. This turns out to be $M_\odot = 2 \times 10^{30}$ kg. The symbol $\odot$ is used to denote any quantity associated with the sun.

Luminosity and surface temperature: Consider the amount of energy radiated by the sun in all directions per second (measured in watts). This is called the solar luminosity $L_\odot$. In order to determine this, we expose a black surface of known area to the radiation from the sun and note the rise in its temperature within a given time. From this the amount of heat received by it can be calculated. This amounts to 2 cal/cm^2.s or 1.388×10^3 W/m^2. This measurable quantity is known as the solar constant. Now, suppose the distance from the sun's centre to the earth is r. Then the amount of radiation received per unit area at the earth's surface per second is simply $L_\odot/4\pi r^2$ (total energy radiated in all directions/area of the sphere of radius r) which is equal to the solar constant. Assuming r to be 1AU (in km), $L_\odot = 3.9 \times 10^{26}$ W. Next, let us assume that this radiation comes from the surface of the sun taken as a sphere of radius $R_\odot$ (696,000 km). The flux emitted by one square metre of sun's surface is given by $E = L_\odot/4\pi R_\odot^2$. Then by the Stefan-Boltzmann law we have $E = \sigma T^4$, where T is the temperature of the sun's surface and $\sigma = 5.669 \times 10^{-8}$ W/m^2.K^4. From these two relations for E we finally get the temperature of the sun's surface to be about 5800 K.

Example 15.6: Assuming that the earth reemits all the radiation it receives from the sun, determine its temperature.

Answer: The amount of heat received by the earth from the sun is 1.388×10^3 W/m^2. If this amount is reemitted, then it can be equated to σT_e^4 where T_e is the temperature of the earth. Therefore we find

$$T_e = \left(\frac{1.388 \times 10^3}{5.669 \times 10^{-8}}\right)^{1/4} = 395\text{K}$$

Solar Interior: The interior of the sun, or of any other star, cannot be observed directly. However, based on the known data - either directly observed or inferred - such as temperature and chemical composition near the surface, luminosity, mass and so on, theoretical models are constructed. These make use of various physial principles and processes such as hydrostatics, thermodynamics, transport of energy and energy generation due to neuclear rections. A great deal of information can be obtained from such models throwing light on the internal structure and properties of stars. For instance, it can be shown that as we go deeper into the interior of the sun, the temperature rises and reaches a value of about 15 milliion kelvin at the centre. We find that the sun is a hot gaseous spehere predominantly made of hydrogen (70%) with other constituents being helium (28%) and other heavier elements (all together amounting to 2%). The average density is about one and a half times that of water. However, the value of the density varies from 10 kg/m^3 at the surface to some 10^4 kg/m^3 at the centre.

Energy production in the sun: 'How does the sun keep shining?' This question intrigued astronomers for a very long time. There must be some source within the sun supplying steadily the energy radiated away by the sun. In the last century two scientists, Helmholtz and Kelvin, suggested that the slow contraction of the sun was responsible for this energy release: the gravitational energy liberated was converted to heat and light. But we can show that this assumption leads to a very short age for the sun. Let us assume that the sun is a uniform sphere. Just as a particle in a gravitational field possesses gravitational potential energy, it can be shown by actual claculation that a sphere of radius R and mass M possesses a gravitational potential energy

$$V = -\frac{3}{5}\frac{GM^2}{R}. \tag{15.5}$$

This is due to the work done in bringing each minute part of the sphere together from infinity to build up the sphere. If the sphere shrinks from infinite radius to radius R the amount of energy released is given by the numerical value of Eq. (15.5). Suppose the sun has shrunk from infinite radius to its present radius R. The total energy released during this process is

$$E = \frac{3}{5}\frac{GM_\odot^2}{R_\odot}. \tag{15.6}$$

Substituting for $M_\odot$ and $R_\odot$, we obtain

$$E = 2.3 \times 10^{41}\,\text{J}. \tag{15.7}$$

As we have seen, the energy radiated by the sun per second is

$$L_\odot = 3.9 \times 10^{26}\,\text{W} \tag{15.8}$$

If the sun has been emitting radiation at this rate since its birth, its age would be

$$t = \frac{E}{L_\odot} \approx 6 \times 10^{14}\,\text{s} \tag{15.9}$$

or about 20 millions years. But the age of the earth has been found by geophysicists to be no less than about 4 billion years. Since t is much smaller than this, the source of solar energy cannot be its gravitational contraction.

After the discovery of radioactivity it was conjectured that the energy from the sun was nuclear in origin. However, it was only in 1939 that Hans Bethe and others showed that the energy production within the sun is thermonuclear in nature by presenting the actual reactions involved. At a temperature of around 10^7K, which is attained at the centre of the sun or any other similar star, four

hydrogen nuclei can combine to form a helium nucleus. The mass difference between four hydrogen nuclei and one helium nucleus is converted ino energy according to the formula $E = mc^2$ thereby achieving the enormous energy release needed to keep the sun shining. We may note that a process similar to above occurs within the man-made hydrogen bomb!

Properties of stars in general: As mentioned earlier, modern astronomers have striven to gather information on different aspects of stars like their luminosities, masses, tempertures, sizes and chemical compositions. In recent years enormous progress has been made in determining the structure and evolution of stars by utilizing this information.

Brightness and luminosity: The brightness of stars is represented through the system of *magnitudes*. This system was initiated as far back as two thousand years ago by the Greek astronomer Hipparchus. He termed the brightest visible stars as first-magnitude stars. Those about one-half as bright were called second magnitude stars and so on to the sixth-magnitude stars which were the faintest stars in the sky. The magnitude scale of Hipparchus was extended with the increasing use of telescopes, since even dimmer stars could be seen through them. In the last century, the techniques of photometry were developed to measure the amount of light received from the stars. This helped in defining the magnitude scale more exactly. Photometric measurements showed that a first-magnitude star is about 100 times as bright as a sixth-magnitude star. The magnitude scale is now defined so that a magnitude difference of 5 corresponds exactly to a factor of 100 in the amount of light energy arriving at the earth. A difference in magnitude by one then corresponds to a difference in the ratio of brightness (or light received) by fifth root of 100 which is 2.512.

The above discussion implies that the relation between any two magnitudes m_1 and m_2 corresponding to brightness values l_1 and l_2 is given by

$$\frac{l_1}{l_2} = 100^{(m_2-m_1)/5} \tag{15.10}$$

Taking the logarithm of the equation on both sides we get, $m_2 - m_1 = -2.5 \log l_2/l_1$. Or in general

$$m = -2.5 \log \frac{l}{l_0} \tag{15.11}$$

Here l_0 is the brightness of a standard star of zero apparent magnitude, since for $l = l_0$ we get $m = 0$. For instance, the star Vega is of zero magnitude and its brightness is $l_0 = 2.52 \times 10^{-8}$ Wm^{-2}. Stars that are brighter than those of zero magnitude have negative magnitude. Thus, a magnitude of -5 corresponds to a brightness 100 times greater than that of a zero magnitude star. We list in Table 15.1 the apparent magnitudes of some characteristic astronomical objects and the limiting values associated with different modes of observation.

Example 15.7: Assuming that the dimmest star visible to the naked eye has a magnitude of about 6, compare its brightness with that of planet Venus.

Answer: The magnitude of Venus is $m_1 = -4$. Magnitude of the dimmest star $m_2 = 6$. Hence $m_2 - m_1 = 10$. Correspondingly, we have the ratio of brightness.

$$\frac{l_1}{l_2} = 100^{(m_2-m_1)/5} = 100^2 = 10,000.$$

Therefore, Venus is 10,000 times brighter than the dimmest visible star.

Table 15.1

Sun	−26.5	Jupiter, Mars	−2
Full moon	−12.5	Sirius	−1.5
Venus	−4	Aldebaran, Altair	1
Naked-eye limit	6.5	5 metre telescope (visual)	20
Binoculars	9	5 metre telescope (photo)	24
20 cm telescope	14		

Stellar spectra: The spectrum of any star, as that of the sun, displays a continuous spectrum interspersed with dark absorption lines. The hot surface layers in which the continuous spectrum is produced is known as the photosphere. The photosphere emits radiation in all wavelengths and the continuous spectrum is surrounded by comparatively cooler layers. The atoms and molecules in these layers absorb radiation of particular wavelengths thereby producing the dark lines. By analysing these lines one can identify the elements and their abundance in the stellar atmosphere.

The continuous spectrum of the star, i.e. the intensity distribution of radiation as a function of wavelengths, decides the apparent colour of the star. One can also determine the temperature of the star from this. This is possible because the stellar spectrum in general has the same form as that of black body radiation. Suppose the maximum intensity occurs at the wavelength λ_{max}. The colour corresponding to the value of λ_{max} decides the colour of the star. For instance if λ_{max} is between 450 nm and 490 nm, the star will appear blue, between 620 nm and 770 nm it will look red and so on. Further, since the spectrum is that of black body radiation we have Wien's law $\lambda_{max}T$ = constant $= 2.897 \times 10^{-3}$ mK where T is the temperature of the emitter of radiation, namely the photosphere of the star. From the spectrum, λ_{max} is observed and hence T can be calculated. Hottest stars appear white or blue, the coolest red or orange and stars with intermediate temperatures yellow or green. Stars are classified into seven principal categories denoted by the letters O, B, A, F, G, K and M. They are in the order of decreasing temperature varying in colour from very blue to red. Our sun is a G star corresponding to $\lambda_m = 550$ nm.

Sizes of stars: Stars appear as pin points of light in the sky owing to their large distances from us. This is true even when they are observed through the most powerful telescopes. The usual method of measuring the angular diameter as is done in the case of planets cannot be applied in the case of the stars. However, the radius of a star can be deduced from its luminosity and temperature. The principle of this method has already been used in finding the temperature of the sun. If the energy emitted by unit area of the star is E, then $E = \sigma T^4$, where T is its surface temperature and σ is the Stefan-Boltzmann constant. If the radius of the star is R, then the luminosity of the star can be expressed as

$$L = 4\pi R^2 E = 4\pi\sigma R^2 T^4. \tag{15.12}$$

In this formula, T is knwon from the stellar spectrum and luminosity L can be obtained from photometric and distance measurements, that is from the measured apparent luminosity L and distance $r (L = 4\pi r^2 l)$. Therefore radius R of the star can be calculated. Majority of stars have radii in the range of a tenth to twenty times the solar

radius. However, there exists 'giants' and 'supergiants', e.g. Capella and Betelgeuse, with radii some 50 to 250 times the size of the sun. White dwarfs, which we shall discuss later, have very small sizes comparable to that of the earth. Common stars of solar size are called 'dwarfs' on this scale.

Masses of stars: We have seen how the mass of the sun can be determined from the motion of the planets around it, and the mass of a planet from the motion of satellites around it. Similarly when two stars form a gravitationally bound system and go round in circles around their common centre of mass, their individual masses can be determined by observing their motion. Such a star system is called a binary star system and nearly half the stars around us are binaries. If the masses of two such stars forming a binary are M_1 and M_2 separated by a and revolving with period T, then Kepler's third law can be shown to yield

$$M_1 + M_2 = \frac{4\pi^2}{G}\frac{a^3}{T^2}. \qquad (15.13)$$

If the masses are measured in terms of solar mass $M_\odot, T$ in years then one can show $G = 4\pi^2$ in such a system of units provided a is expressed in AU. So we can write, in these units,

$$M_1 + M_2 = \frac{a^3}{T^2}. \qquad (15.14)$$

If the centre of mass can be located, then the distances a_1 and a_2 from it to M_1 and M_2 respectively can be determined. With the help of Eq. (15.14) and the relations $a_1 + a_2 = a$ and $M_1 a_1 = M_2 a_2$, M_1 and M_2 can be obtained individually. By this method it has been determined that the components of Sirius have masses of about $1 M_\odot$ and 2.4 $M_\odot$

It is found in general that the intrinsic brightness or the luminosity of stars increases with mass. As a matter of fact L varies as M^α, where α varies from 2 to 4. For stars similar to the sun $\alpha \approx 4$. Thus stars about forty times heavier than the sun are roughly million times more luminous than the sun.

Stellar interiors: Knowing the mass and the size of stars, we can compute their mean densities. The density varies from some 50,000 kg/m^3 in the case of the coolest M stars to around 10 kg/m^3 for the hot O stars. Within the star, density and temperature increase towards the centre. Central densities range from a few thousand kg/m^3 (O stars) to about 10 kg/m^3 (M stars). Similarly the central temperature ranges from 10 to 30 million degrees as we go from M to O stars.

Stellar evolution: As we have mentioned, every star has its own life consisting of different stages of birth, evolution and death. In the vast spaces between stars there exist enormous clouds of gas and dust. Orion Nebula, the nearest such cloud, can be faintly seen in the constellation Orion and offers a spectacular sight when veiwed through a powerful telescope. Often large clumps form within such a cloud and progressively contract under their own internal gravitational fields. These clumps may further fragment and each of these fragments can grow attracting more mass, contracting and heating up at the same time. When the temperature rises high enough, nuclear reactions can be triggered. The result is the birth of stars from the original gas cloud. The gravitational pull towards the centre of the star due to its mass is balanced by the outward pressure due to the heat generated by the nuclear reaction. This leads to a stable configuration such as that of the sun in its present state. As was discussed earlier, the energy production inside the sun occurs by the conversion of hydrogen into helium. The sun was born

some four and a half billion years ago. The nuclear burning can continue until all the hydrogen in the centre of the sun is converted into helium. This would happen in another five billion years signalling the death of the sun. If the stars are more massive than the sun, more complicated nuclear reactions can take place leading to the production of elements of higher atomic numbers. Accordingly, the lifespan and the stages of evolution depend on the mass of the star..

Returning to the sun, when the central core is totally converted into helium it can no longer sustain the heat production and the core will contract. This will happen roughtly in another five billion years. The outer part bloats up and becomes cooler thereby assuming a red hue. The sun will have become a red giant. The size of the sun will be so enormous that Mercury and Venus will be swallowed up and the earth burnt in the heat.

The core of a star that has exhausted its nuclear fuel can end up in three ways depending on its mass. In the case of a star like the sun, the gravitational compression leaves the core composed of protons with electrons flying around forming a gas-like phase - the electron gas. This electron gas can withstand the inward gravitational force and a stable equilibrium would be achieved. Such a star is called a white dwarf. It gradually cools, ceases to be self luminuous and ends up as a black dwarf.

It was theoretically discovered by S. Chandrasekhar (1910-1995) in the 1930's that the above stable configuration could occur only for cores of mass up to 1.4 $M_\odot$. This is known as the Chandrasekar limit, Beyond this mass the gravitational compression becomes so large that the electrons are forced into the nuclei, a process that is called inverse beta decay. The star is now composed of only neutrons that can withstand the gravitational compression. We now have a neutral star. The stellar matter is of nuclear density. Once it was unknown how one could detect a neutron star which would be nonluminous. However, now it is firmly believed that pulsars (Pulsating Radio Sources) discovered in 1867 are indeed neutron stars. Pulsars are endowed with enormous magnetic fields of some 10^{12} Gauss and beam radiation in a narrow cone. As the pulsar rotates rapidly, the beam crosses the observer at regular intervals giving rise to closely spaced pulses. Pulsars with periods from millisceonds to several seconds have been detected. During the formation of a neutron star by the contraction of the central part of the star, enormous amount of gravitational energy is released. Because of this energy release the outer layers of the star explode. The star's brightness increases sharply for sometime and then diminishes. This phenomenon is knwon as the supernova.

Even the neutron star has a mass limit. This is definitely less than around 5 $M_\odot$. Beyond this mass, the gravitational inward force will be so large that even the neutrons forming a gas cannot withstand it. Then the core undergoes catastrophic gravitational contraction which no known force in nature can stop. Once the star of mass M has contracted within a radium $r_{BH} = 2GM/c^2$, a black hole is said to have been formed. The gravitational pull within this radius is so strong nothing, not even light, can escape this region. Hence the name 'black hole'. The stellar matter keeps contracting, the entire mass getting compactified towards the centre with the density tending to infinity. Nobody knows at present what exactly happens to this mass in its ultimate state of compression. For a far off outside observer, however, the black hole is represented by space with the peculiar property that any thing entering r_{BH} being swallowed up while outside this radius one would

detect normal gravitational pull due to a star of mass M.

We have so far briefly outlined the properties of stars and their evolution. We shall now move on to the large scale structure of the universe as a whole.

15.4 The Milky Way

On a clear night we can see among the stars a hazy, luminous band stretched across the sky. This is the 'Milky Way' or Akashaganga, a celestial entity which once inspired a variety of mythological explanations in different cultures. It is now well known that this band of diffuse glow is in fact composed of a few hundred billion stars including our own sun. The study of the Milky Way, or the galaxy to which we belong, signifies the first step in the exploration of the large scale structure of the universe.

A systematic study of the Milky Way was undertaken in the eighteenth century by William Herschel, the same Herschel who discovered the planet Uranus. By observing the Milky Way through a large telescope of 48 inch diameter he concluded that the hazy band was composed of at least a million stars and that the solar system belonged to this enormous system of stars. By examining the distribution of stars in all directions he further surmised that the sun is at the centre of the Milky Way.

The true picture of the Milky Way emerged only in the present century beginning with the work of Harlow Shapley. He examined the distribution of star systems called globular clusters that contain 10^5 to 10^6 stars. From their distribution and their distances from us he could roughly construct a model for our galaxy. He also discovered that the sun is not at the centre of the Milky Way as was orginally supposed.

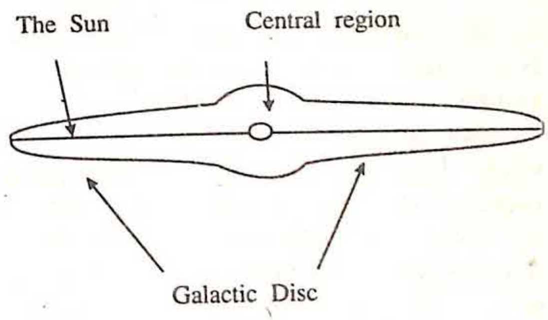


Figure 15.4: Schematic diagram of the galaxy.

Today because of the continued observations involving both visible and radio waves, we know a good deal about our galaxy. Viewed edge on it looks like a disk with a central bulge; the density of stars falling off away from the centre. The diameter of this enormous conglomorate of stars is some 100,000 light years. The sun is situated about two thirds distance away from the centre. The cental thickness of the Milky Way is about 5000 light years and the thickness in the solar neighbourhood is around 1000 light years. The structure of the Milky Way has been investigated by observations involving radiowaves coming from different regions. A characteristic feature revealed by these observations are the spiral arms emanating from the inner regions of the Milky Way.

The Milky Way comprises not only individuals stars but also star clusters. 'Open clusters' consist of well separated stars ranging in number from 100 to 1000. As we have seen, there are also globular clusters, so named because of their overall spherical distribution, containing much larger numbers of stars. Interstellar space is also interspersed with gas and dust clouds. Some regions of the Milky Way appear dark not because they are empty but because stars in those regions are obscured by intervening gas and dust.

A remarkable feature of the Milky Way is

its rotation about its centre. This rotation is not rigid. Each individual star moves in an orbit with speed determined by the gravitational pull of the stars enclosed within its orbit. Thus our sun, along with the planetary system, is revolving around the centre of the Milky Way with a speed $v = 250$ km/s. It takes about 250 million years for the sun to go round once. The last time the sun was at its present position, dinosaurs roamed the earth!

The mass of the Milky Way can be roughly estimated knowing the distance of the sun d from the centre, and its orbital speed v. Assuming the gravitational pull is due to a mass M effectively concentrated at the centre, we have

$$\frac{GM}{d^2} = \frac{v^2}{d} \text{ or } M = \frac{v^2 d}{G}. \qquad (15.15)$$

Setting $v = 2.5 \times 10^2$ m/s, $d = 3 \times 10^4$ ly $\simeq 3 \times 10^{20}$m, we obtain $M = 3 \times 10^{41}$ kg. This is about 150 billion solar masses. Therefore we may conclude that the part of our Milky Way within the orbit of the sun is made of around 150 billion stars. If we include the stars exterior to this orbit, the total number of stars in the Milky Way would be higher than this estimate. We see that we live in a galaxy, our Milky Way, which is an immense collection of stars. However, as we shall see our galaxy is merely a single cosmic building block. It takes billions of such building blocks to make up the universe at large.

15.5 The structure and evolution of the universe

Our galaxy, the Milky Way, is so large with so much variety that one could easily feel that it is the universe by itself. This is not so. Two centuries ago the philosopher Immanuel Kant (1744- 1804) had conjectured that some of the faint luminous patches in the sky, or the nebulae as they were called, were in fact distant systems of stars comparable to our Milky Way. He called these systems, including the Milky Way, 'island universes' - enormous isolated communities of stars populating an immense empty space. It was Edwin Huble who in the 1920's established the truth of this conjecture.

Edwin Hubble pursued a systematic study of the 'nebulae' which we know today to be actually galaxies like our own. He not only observed, photographed and classified the galaxies, but also measured the distances to them. The distance to the closest one in the constellation Andromeda turned out to be about two million light years. Since the size of the Milky Way is 10^5 light years, this clearly established that one was observing a celestial object that was outside of the Milky Way. Further, its distance and the angle subtended showed that this celestial object was comparable in size to the Milky Way. It was rightly concluded that it was indeed a galaxy like our own. Hubble went on to measure distances to galaxies much farther than the above one. Since Hubble's days astronomers have gathered a wealth of information on the distribution, distances and structures of galaxies. On large distance scales of 10^8 - 10^9ly, the distribution of the galaxies seem to be quite uniform.

This means that we are living in a homogeneous universe in which all points are alike. Copernicus showed that there was nothing special about the earth which, after all, was not at the centre of the solar system. The solar system was itself shown to be far from the centre of our Milky Way galaxy. We see now that our galaxy enjoys no special status since all parts of the universe are similar. This idea of the sameness of the constituent parts of the universe is known as the Copernican Principle. Ours is but one of an awesome number of galaxies all of which should be given equal status. Within the visible range

of some 10 billion light years, accessible only to the most powerful telescopes on earth, one can see aroung 10 billion galaxies. Thus even the visible part of the universe is unimaginably immense. And the galaxies are the building blocks of our vast universe.

We have very briefly touched upon the spatial structure of the universe. But what about its behaviour in time? Since the time of the philosopher Aristotle, it was believed that the universe was static and unchanging on a large scale. This was shown to be false in the 1920's thereby leading to the realization that we are living in a dynamic, evolving universe.

When studying the spectra of different galaxies a peculiar phenomenon was observed. Characterisic spectral lines produced by elements present in any celestial source can be identified by comparison with the spectra of elements observed in earth based laboratories. In the case of any given galaxy the spectra was observed to be shifted towards the red. The spectral shift is measured by the quantity.

$$z = \frac{\lambda - \lambda_0}{\lambda_0}, \tag{15.16}$$

where λ_0 is the emitted wavelength (the same as in the terrestrial laboratory) and λ is the observed wavelength. A positive $z(\lambda > \lambda_0)$ gives a 'red shift'. All galaxies invariably showed red shift of spectral lines which was the same for all lines of any particular galaxy. The red shift vaired from galaxy to galaxy.

As we have seen earlier in this book, shift in observed wavelength or frequency was first discovered in the case of sound waves by the Dutch scientist Doppler. We have seen that this phenomenon occurs in the case of light waves also (Chapter 10). The shift z, if small, is given by the speed of the source divided by the speed of the waves, being c in the case of light. Therefore the red shifts of galaxies can be written as $z \simeq v/c$ where v is the speed with which the galaxy is moving away. The red shifts of galaxies showed that they were rushing away from us with enormous speeds. This showed that the universe built up by the galaxies was far from being a static one. Indeed it demonstrated that we were living in an expanding universe.

Example 15.8: What is known as the K line of singly ionized calcium has a wavelength of 393.3 nm as measured on earth. In the spectrum of one of the observed galaxies, this spectral line is located at 401.8 nm. Determine the speed with which this galaxy is moving away from us.

Answer: The red shift of the galaxy is given by

$$z = \frac{401.8 - 393.3}{393.3} = 0.0216.$$

The galaxy is then moving away from us with a speed

$$\begin{aligned} v &= zc = (0.0216).(3 \times 10^5 \text{km/s}) \\ &= 6480\,\text{km/s}. \end{aligned}$$

Edwin Hubble and his coworker Milton Humason made a systematic study of the red shifts of various galaxies. They made the remarkable discovery that the red shift z is directly proportional to the distance r of a galaxy from us. Since $z = v/c$, this means the speed of recession of a galaxy is proportional to its distance r. This is expressed as

$$v = Hr, \tag{15.17}$$

which is known as Hubble's law. Let us note that $(1/H)$ has the dimension of time; we shall see its significance presently. If the galaxies are rushing away from one another,

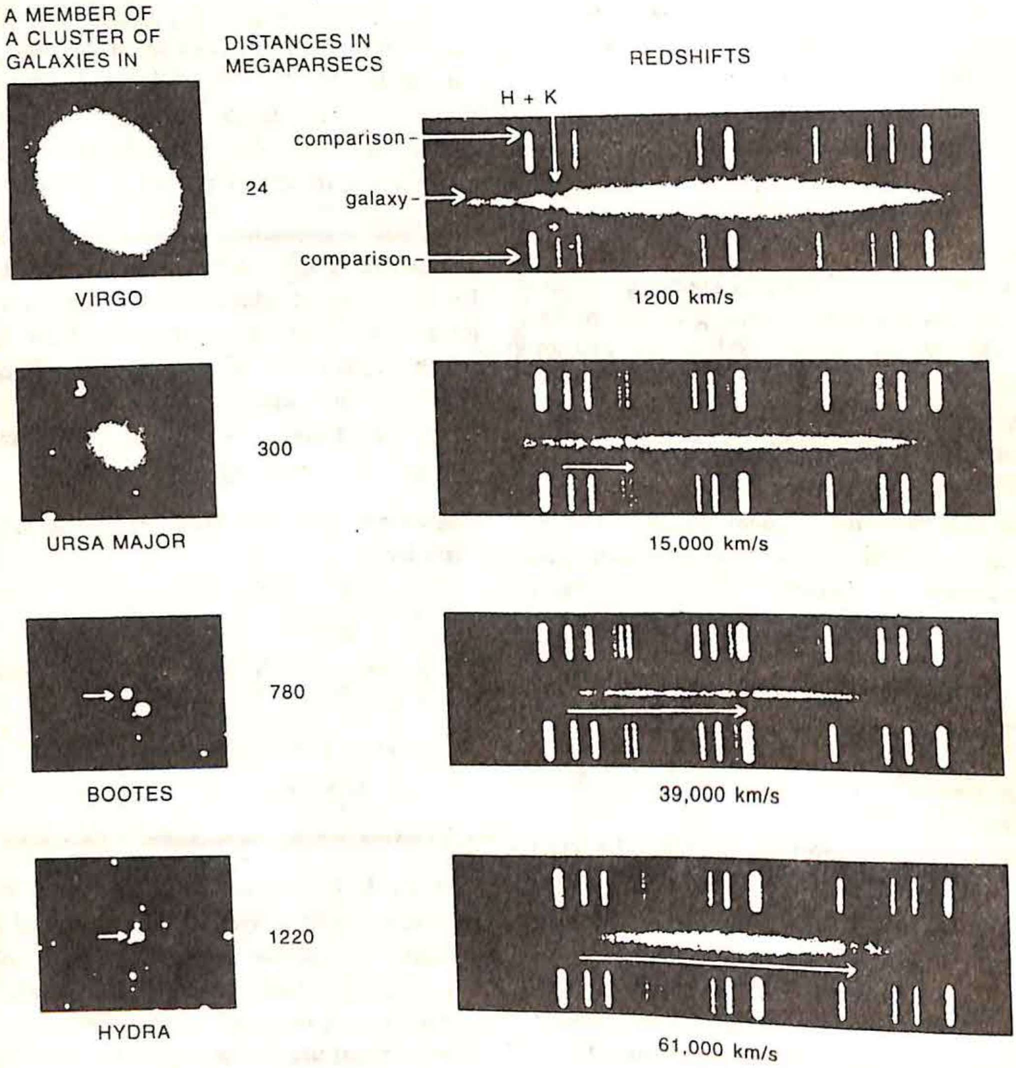


Figure 15.5: Galaxies and their spectra. The photographs of the galaxies have the same magnification and are arranged in order of increasing distance from the earth. The spectrum of the galaxy is the hazy band between the two comparison spectra. The horizontal arrow below the spectrum of each galaxy points to the shifted H and K lines of singly ionized calcium. As we go down the figure, the length of the horizontal arrow increases. This means that for more distant galaxies, the shift of the spectral lines to longer wavelengths (red shift) is greater.

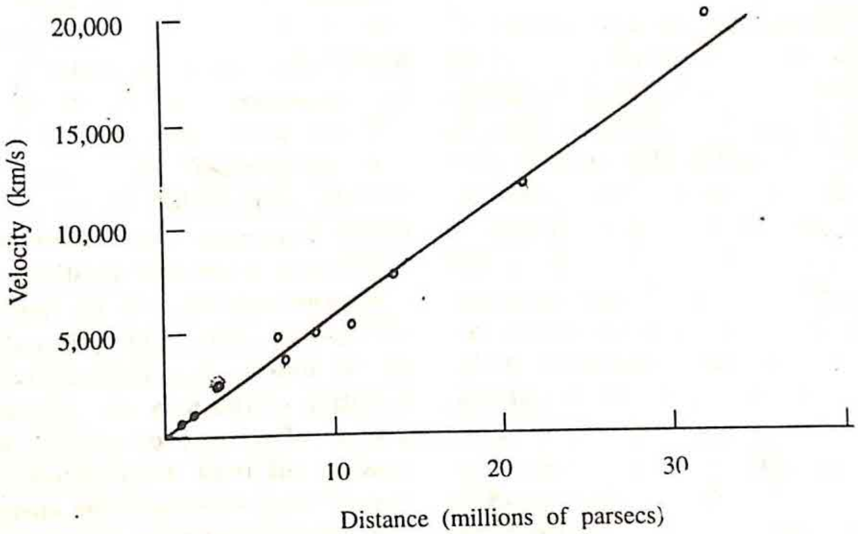


Figure 15.6: Hubble's velocity-distance relation (shown by the straight line). Circles represent the experimentally observed values for some galaxies. (1 parsec = 3.26 light years)

then if we go backwards in time, we see that they must have been closer together at an earlier epoch. Then the question arises whether we can arrive at a moment in the remote past when all the galaxies were together, a moment which represents the beginning of the universe itself. Hubble's law indicates that this is in fact true. Suppose all galaxies including ours burst out from a single point at time $t = 0$ with constant speeds with respect to us given by $v_1, v_2, \cdots$ etc. They would all travel in time t_0 distances $r_1 = t_0 v_1, r_2 = t_0 v_2 \cdots$. This shows in general that $r = (1/t_0)v$ which has the same form as Hubble's law. Therefore we can identify $(1/H)$ with t_0, the time that has elapsed since the galaxies exploded from a common point or at least a common neighbourhood. The quantity t_0 would then be the age of the universe. Since all galaxies including ours started in a similar way there is nothing special about our location although we make calculations with respect to our own frame of reference. Therefore it is not that all other galaxies are moving away from us as the cenral point. Galaxies are simply moving away from one another and the picture would be the same from any given galaxy. The quantity $t_0 = (1/H)$, which gives us an idea of the age of the universe, is estimated to be around 10 to 20 billion years. When we go back in time to this era, we reach the very beginning of the universe. At this time, galaxies and stars had no individual existence. All matter and radiation were compressed into a fiery region with extremely high temperature and immense densities. This is known as the 'primordial fireball'. The awesome explosion, which launched the expansion of the universe, is called the 'Big Bang'. As matter and radiation cooled, stars and galaxies formed, planets came into existence and finally, at least on the earth, life emerged. All

this as we have pointed out has taken billions of years.

Compelling observational evidence for the Big Bang theory of the universe came in the sixties. If the universe has been expanding since its beginning, the radiation once extraordinarily hot within the 'fireball' must be cooling steadily just as a gas cools when its enclosure is expanded. If the radiation has a black body spectrum the wavelength at which maximum intensity occurs is correlated with temperature T by Wien's law $\lambda_{max}T$ = constant. The temperture of this radiation spread out over the entire universe can be shown to be about 3K. Such an all pervading radiation was in fact detected by Arno Penzias and Robert Wilson in 1965. This gave excellent support to the Big Bang model of the universe.

What about the future of the universe? There are two courses open to its further evolution. To understand this, consider the analogy of a number of stones thrown out from the surface of the earth. If there speeds are high enough to effectively overcome the gravitational pull of the earth, the stones will fly away never returning to earth. On the other hand, if the speeds are not sufficient, the gravitational pull can prove strong enough to bring first the stones to a momentary standstill after which the stones will fall back to earth. On the cosmic level the gravitational attraction is provided by the entire mass-energy content of the universe. If this falls short of a critical value, the gravitational pull cannot overcome the speeding galaxies which will keep flying forever. The universe will expand forever in this case, stars, galaxies and all else slowly dying out. On the other hand, if the mass energy content is above a critical value, the galaxies will slow down essentially under the influence of their own cumulative gavity. They will come to a standstill and then start moving towards one another. Equivalently, the universe will expand to a minimum limit and then contract. It will reach the primordial fireball state as in the beginning and it has been speculated that it will explode with a Big Bang again. We do not know how much mass-energy there is in the universe. Therefore the ultimate fate of our universe is unknown at present. Only future observations will have to settle this question.

As was mentioned at the very beginning of this book, gravitation plays a paramount role in determining the structure and the evolution of the universe. To describe the interior of a star, of course, we have to consider all basic interactions: weak and strong forces enter into the energy production, electromagnetism plays an important role through charge interactions and radiation fields, while gravitation holds the star together. But a star on the whole is electrically neutral and over long ranges gravitation is the only effective force. Consequently, the internal dynamics of galaxies composed of stars and the large scale structure of the universe made up of galaxies are both controlled essentially by gravitation. On cosmic scales, therefore, the weakest force in nature is the supreme ruler!

Although we have an excellent idea of the large scale structure and evolution of the universe, there are a number of challenging questions awaiting proper answers. For instance, we do not know exactly how galaxies were formed and the details of the formation of their structure. Equally obscure is the reason for the present uniformity of the universe. Has the universe always been homogeneous or did it evolve into its present state from an earlier chaotic condition? At a much more fundamental level, we would like to understand more clearly the very creation of the universe. Physicists have been trying to describe the processes taking place

extremely close to that moment. Although, as we have remarked, gravitation decides the cosmic structure at later stages, all the four fundamental forces become equally important at the very early times. As a mater of fact, in order to probe the universe near the moment of creation we need a theory, as yet unformulated, that unites all these forces including gravitation.

Cosmology, the science that tries to describe the universe, began ages ago when man first wondered about the mysterious heavens around him. The immense journey of exploring the universe that started then will continue as long as man does not lose that sense of wonder and curiosity.

Summary

1. In the geocentric model, the earth is stationary and all celestial objects revolve round it. Each planet goes round in a small circle (epicycle), the centre of which follows a circular path round the earth. With improved observations, the model became complicated with increasing number of epicycles added at each stage. In the much simpler heliocentric theory, the sun is at rest and the planets revolve round it. The three laws of planetary motion due to Kepler are explained by Newton's universal law of gravitation.

2. *Planets.*

Distance: In the parallax method, distance D of a planet from the earth is determined by $\theta = b/D$ where θ is the angle between the two directions of viewing the planet at the same time from two points on the earth a distance b apart.

The angle between the earth-planet direction and earth-sun direction is called the planet's elongation. When this angle attains its maximum value ϵ, (see Fig.15.3).

$$r_{ps} = r_{es}\sin\epsilon; \; r_{pe} = r_{es}\cos\epsilon$$

Knowing ϵ and r_{pe}, r_{ps} and r_{es} can be determined.

Knowing the distance of any one planet from the sun, that of any other can be calculated using Kepler's third law:

$$\frac{a_2^3}{a_1^3} = \frac{T_2^2}{T_1^2}$$

where the periods of revolution T_1 and T_2 are determined by direct observation.

The average earth-sun distance is called the astronomical unit (AU). This can be determined by the parallax method. 1 AU = 1.496×11^{11} m. Mean distances of planets range from 0.39 AU (Mercury) to 39.46 AU (Pluto).

Size: If d is the diameter of a planet and α is the angle subtended by d at the earth, then $\alpha = d/D$. Measuring α, the size of the planet is determined. d ranges from 3000 km (Pluto) to 1,43,000 km (Jupiter). The same method gives the diameter of the visible part of the Sun to be about 1,400,000 km.

Mass: If a planet of mass M has a satellite in a orbit of radius R and period T,

$$M = \frac{4\pi^2 R^3}{GT^2}$$

Knowing R and T, M can be determined. In terms of the mass of the earth, planetary masses range from 0.002 (Pluto) to 318 (Jupiter).

The same formula gives mass of the Sun $M_\odot$ to be about 2×10^{30} kg. In this case R and T refer to the motion of a planet around the sun.

Atmosphere: Of a planet is determined by its mass and temperature. The chemical composition is ascertained from spectrographic analysis of sunlight reflected from the planet. Other constituents of the solar system are comets moving in highly elongated elliptic orbits, asteroids or minor planets and meteoroids.

3. *The Sun*: Distance, size and mass of the sun are given above. Solar luminosity $L_\odot$ is the amount of energy radiated by the sun in all directions per second. To determine $L_\odot$, we measure the amount of radiant energy per unit area per second at the earth's surface, which equals $L_\odot/4\pi r^2$ where r is the sun-earth distance. $L_\odot = 3.9 \times 10^{26}$W.

   From $L_\odot = 4\pi\sigma R_\odot^2 T^4$ where $R_\odot$ is the radius of the sun and σ is the Stefan-Boltzmann constant, the surface temperature T of the sun is found to be about 5800 K. At the centre of the sun, temperature is estimated to reach about 15 million K. The sun is made of hydrogen (70%), helium (28%) and other heavy elements (2%). If one assumes that radiant energy from the sun arises due to the gravitational energy released in its slow contraction, the age of the sun would be about 20 million years, which is in conflict with the geophysical estimate of the earth's age ($\sim$ 4 billion years). It is now established that energy production in the sun arises from nuclear fusion reactions (4 hydrogen nuclei fusing into one helium nucleus) at the high temperature of its core.

4. *The Stars*: (in general)

   Brightness ℓ of a star is measured through the system of *magnitudes* defined by

$$m = -2.5 \log \frac{\ell}{\ell_0}$$

or

$$m_2 - m_1 = -2.5 \log \frac{\ell_2}{\ell_1}$$

where ℓ_0 is the brightness of a standard star of zero magnitude. For example, Venus has magnitude -4. A magnitude of -5 corresponds to a brightness 100 times greater than that of a zero magnitude star.

Spectra: The hot surface layers of a star (photosphere) emit continuous spectrum of radiation. The relatively cooler layers surrounding the photosphere absorb radiations of particular wavelengths depending upon composition, and produce dark lines in the spectrum. If the maximum intensity occurs at wavelength λ_{max}, Wien's law; $\lambda_{max}T = 2.897 \times 10^{-3}$ mK can be used to find the temperature of the photosphere.

Size: The radius R of a star is determined by the relations. $L = 4\pi\sigma R^2 T^4$ and $L = 4\pi r^2 \ell$. L is luminosity; ℓ is apparent luminosity obtained from photometric measurements, r is the distance of the star and T is its surface temperature obtained from the steller spectrum.

Mass: If two stars forming a binary (masses M_1 and M_2) are separated by 'a' and revolving around their common centre of mass with period T,

$$M_1 + M_2 = \frac{4\pi^2}{G}\frac{a^3}{T^2}$$

Also, $M_1 a_1 = M_2 a_2, a_1 + a_2 = a$

where a_1 and a_2 are the distances of M_1 and M_2 from their centre of mass. From these relations M_1 and M_2 can be determined.

Stellar evolution: In a stable configuration, the inward gravitational pull is balanced by the outward pressure due to the heat generated by nuclear burning of hydrogen. When the fuel is exhausted, star cores of mass upto 1.4 $M_\odot$ (Chandrasekhar limit) end up as white dwarfs in which the electron gas withstands gravitational pull. For greater masses, they end up as neutron stars (e.g. pulsars). Beyond about 5 $M_\odot$, they end up as black holes, from which nothing, not even light, can escape.

5. *The Milky way*: is our galaxy containing billions of stars. Its diameter is about 10^5 light years; the sun is situated two thirds distance away from the centre and revolves around it with a speed 250 km/s. Spiral arms emanate from the inner regions of the Milky way.

6. *The Universe*: contains billions of galaxies. On large distance scales of 10^8 - 10^9 light years, the universe is homogeneous. From measurements of red shifts, the recessional speeds (v) of galaxies can be determined. Hubble discovered that our universe is expanding; $v = Hr$ (Hubble's law) where $1/H$ is around 10 to 20 billion years. The explosion which triggered the expansion of the universe is called the 'Big Bang'. Compelling evidence for the Bing Bang theory came with the discovery of 3K radiation pervading the universe.

Exercises

15.1 When the planet Jupiter is at a distance of 824.7 million kilometers from the earth, its angular diameter is measured to be $35.72''$ of arc. Calculate the diameter of Jupiter.

15.2 Assuming that the orbit of the planet Mercury around the sun to be a circle, Copernicus determine the orbital radius to be 0.38 AU. From this, determined the angle of maximum elongation for Mercury and its distance from the earth when the elongation is maximum.

15.3 Suppose there existed a planet that went around the sun twice as fast as the earth. What would be its orbital size as compared to that of the earth?

15.4 Io, one of the satellites of Jupiter, has an orbital period of 1.769 days and the radius of the orbit is 4.22 $\times$ 10^8 m. Show that the mass of Jupiter is about one thousandth that of the sun.

15.5 Suppose the sun was located at the position occupied by the star nearest to us, namely α-Centauri, which is at a distance of about 4 ly from us. By what factor would the radiation received per unit area per second at the earth be reduced?

15.6 There are certain types of stars called 'variable stars' which undergo periodic change in their light output. If such a star doubles its light output, how much does its magnitude change?

15.7 The phenomenon of a 'nova' involves the sudden outburst of a star. The star then becomes much brighter than usual for a few days or weeks. In 1975, a nova appeared in the constellation of Cygnus (the Swan). In two days the magnitude of the stars changed from +15 to +2. By what factor did its brightness increases?

15.8 Work out the ranges of temperature corresponding to which a star will appear blue and red, respectively.

15.9 Consider a binary star system consisting of two stars of masses M_1 and M_2 separated by a distance of 30 AU with a period of revolution equal to 30 years. If one of the two stars is 5 times farther from the centre of mass than the other, show that the masses of the two stars are 5 and 25 times that of the sun.

Additional Exercises

15.10 It is a well-known fact that during a total solar eclipse the disk of the moon almost completely covers the disk of the sun. From this fact and from the information you can gather from examples 15.1 and 15.2. determine the approximate diameter of the moon.

15.11 Suppose the sun shrank from its present size so that its radius is

halved. What would be the change in its gravitational potential energy? (Calculate the actual number in joules).

15.12 In the constellation Orion there is a bright, reddish star called Betelgeuse. Its luminosity is 10,000 times that of the sun and its surface temperature about 3000 K. How much larger is the radius of Betelgeuse compared to that of the sun.

15.13 Consider a white dwarf and a neutron star each of one solar mass. The radius of the white dwarf is that of the earth (about 6,400 km) and the radius of the neutron star is 10 km. Determine the densities of these two types of stars.

15.14 Let us assume that our galaxy consists of 2.5×10^{11} stars each of one solar mass. How long will a star at a distance of 50,000 ly from the galactic centre take to complete one revolution?

15.15 Consider the galaxy mentioned in example 15.8. If it is at a distance of 430 million light years from us, determine Hubble's constant H and the corresponding age of the universe t_0. Note: Express H in $(\text{km/s})/10^6\text{lv}$ and t_0 in years.

Edwin Hubble (1889-1953) American astronomer, orginially trained in law, Hubble was also good at athletics and boxing. His indication of the existence of exchange of extragalactic nebulae initiated the study of the universe beyond our galaxy. He classified galaxies according to their shapes and discussed their evolution. He made the important discovery that the radial velocities of receding galaxies are proportional to their distances from us which is known as the Hubble Law. This has led to the concept of the big bang origin of the universe and its subsequent expansion.

Epilogue

An overview of this book: Electromagnetism as a fundamental force

We have studied many aspects of electricity and magnetism in this book. Let us see how it fits in with other things you have learnt and other parts of physics. In Vol. I you have learnt about Newton's Laws of Motion. There are two important ideas involved here. One is that of *inertial frames of reference* - these are reference frames in which Newton's First Law of Motions is obeyed. (remenber that a reference frame is a choice of an origin and three mutually perpendicular coordinate axes in space, so that we can assign definite valus of cartesian coordinates and times to moving bodies and events that we are interested in). Thus in any such frame, an isolated material body on which no forces act (a condition which can be realised as accurately as we wish if this body is sufficiently far away from all other bodies) travels uniformly in a straight line. The other important idea is that of force - this is what causes *accelerated or nonuniform* motion of a body in an inertial frame. If in such a frame a body is seen to move nonuniformly, there must be some force acting on it.

Many kinds of forces are known and are familiar to us. Some, such as friction, viscosity, drag due to air etc. are not basic or fundamental; they are rather complicated effects or results of other simpler and more fundamental forces. A really fundamental force in nature is the *gravitational force*; its basic statement is contained in *Newton's Law of Universal Gravitation*. As you will remember, this law says that any two bodies with masses m_1 and m_2 separated by a distance r, attract each other with a force having the magnitude

$$F = \frac{G.m_1.m_2}{r^2},$$

and acting in the direction of the straight line joining them. (Here we must suppose that the sizes of the two bodies are negligible compared to the distance separating them, and in that case we can speak of them as point masses). In this Law, if we express the masses m_1 and m_2 in kg, the distance r in m, and the force F in Newtons (N), then the constant G called Newton's gravitational constant has the experimentally measured value

$$G \simeq 6.7 \times 10^{-11} \mathrm{Nm}^2/\mathrm{kg}^2.$$

This is a particular fundamental force in nature. It acts between two bodies however far apart they may be, though it gets weaker as their separation increases. It is the force that holds you and me to the earth, the moon to the earth, the earth and other planets to the sun, and so on to large and larger scales of distance (of course Newton's original description of it based on the assumptions that 3 - D physical space is Euclidean, and that force of gravity acts instantaneously however large the distance separating bodies, has been greatly modified in later times, in the General Theory of Relativity due to Albert Einstein).

There are other fundamental forces in nature besides the force of gravitation. Some of them act between certain material particles only at extemely short distances. They are important only in connection with understanding the "ultimate" structure and properties of matter, i.e. the nature of the nucleus within the atom. (As you would know from your studies of general science and

chemistry, linear dimensions of atoms and nuclei are approximately 10^{-10}m and 10^{-15}m respectively). These are the forces among protons, neutrons, electrons and other subatomic particles; there are two different forces involved here, one called the strong nuclear force and the other the weak nuclear force. It should not be a surprise that there should be more than one kind of force between certain particles; after all, we know that if we have two charged bodies, then both a gravitational and an electrostatic force acts among protons and neutrons only if they are within a distance of 10^{-15} m from one another; beyond this distance this force is quite negligible. The weak nuclear force is present among protons, neutrons, electrons and some other particles as well, but now the "range" or distance upto which it acts is only about 10^{-17}m! Strictly speaking, even though we have used the word "force" in these cases, we can not imagine them to be what appears in Newtons' second law of motion; instead we must use them in a new equation of motion in the framework of what is called Quantum Mechanics. Unlike Newton's Laws of Motion which are the foundation of classical mechanics - this is what we can use to understand the motion of matter in bulk, as we have seen in many examples in Volume I - we need to use quantum mechanics to understand atoms, nuclei and their constituents (as well, in fact, of many properties of matter which have their origin in micorscopic phenomena). In any case, the important point to appreciate is that unlike Newtonian gravity, the ranges of nuclear froces are extremely small. Some aspects of cuclear forces were touched upon when we learned about atoms and nuclei.

There is only one other fundamental force in nature which, like gravitation, acts between certain bodies however far apart they may be (though, just like gravity, it too gets weaker and weaker as the separationkeeps increasing). This is the *electromagnetic force*. We have studied many aspects of this force in this book, but in this summary there is no harm in repeating some of the things we have already learnt. For a long time people thought that there were two different kinds of forces involved here namely *electric forces* and *magnetic forces*; but starting from about the first quater of the 19th century it was gradually understood that these are actually two aspects of one fundamental electromagnetic force. (This is an example of "unification", just as Newton's Law of Universal Gravitation showed that the force of gravity on objects on the surface of the earth the same origin as the force the earth exerts on the moon; terrestrial and celestial gravitation are two examples of a common fundamental law). The electromagnetic force is truly a fundamental force; it cannot be derived from anything more fundamental. In contrast many forces we experience daily and which we mentioned earlier, such as friction, viscosity, van der Waals forces in a gas or liquid etc, are (of course rather complicated) consequences of the basic electromagnetic force.

In the remainder of this Epilogue, we shall recall the bare qualitative aspects of electromagnetic forces. The quantatative treatment has already been given in the previous chapters. Sometimes we will recall historical facts, sometimes jump history. It is to help youto understand what we have already studied as being part of a general picture that this Epilogue has been written.

We have seen that electromagnetic forces act between bodies which are electrically charged, and that these forces are present whatever be the distance separating such bodies, though the force gets weaker as the distance increases. (In this respect it is like force of gravity). But to understand the ef-

fect of electromagnetic force on the motions of bodies on which they act, we can use Newton's Laws of Motion only if these motions are on a macroscopic scale; to find out the effects of electromagnetic forces at atomic or nuclear scales we must use the equations of motion provided by quantum mechanics (as we must in the case of muclear forces as well). For the most part of this Volume, we have happily been dealing only with the former kinds of situations; so then electromagnetic forces (like any other force) are just what can appear "on the right hand side" of Newton's equations of motion for bodies on which such force can act! The electromagnetic force is an example of a force as conceived in Newton's Laws of Motion, occuring in certain situations.

What is the nature of the electric charge on bodies experiencing electromagnetic forces? Go back for a moment to the gravitational case: this force acts between any two bodies with mass, the magnitude being given by Newton's Law of Gravitation. Thus the mass of a body plays, in terms of gravitational force, 2 roles: (i) it acts as the cause or source of a force on any other massive body; (ii) it determines also the strength of the gravitational force experieced by the given body due to other masses in its surroundings. We can think of these as the active and the passive roles of mass respectively. (Of course the mass of the given body plays a special third role: it appears "on the left hand side" of Newton's equation of motion and so determines the acceleration of the body, whatever the force acting on it may be). For electric and magnetic forces, these two roles of mass are played as we have seen by the *electric charge* of a body. While every material body has a positive mass, the elecric charge on a given body can however be nonzero or zero. In the former case, we say the body is *electrically* neutral. Only *electrically charged bodies feel and exert electromagnetic froces on one another.* (To be precise we must include bodies which have distributions of charge but are overall neutral). When we studied the nature of magnetic forces it may at first have seemed that electrically uncharged bodies can feel and produce magnetic forces. But leaving aside some subtle quantum mechanical aspects, we saw that magnetic forces too are present only among electrically charged bodies. In a nutshell the relation between electric and magnetic forces is this: while electric forces exist among charged bodies whether they are in motion or are stationaly, magnetic forces arise only among charged bodies in motion (as seen in an inertial frame).

A nonzero electric charge can be of two kinds, namely *positive* or *negative.* (In contrast the mass a body is always positive). Thus one can have a body with total charge zero, i.e. it is neutral, but which has positive charges somewhere and negative charge elsewhere. In any event, the electric charge of a charged body plays the same role with respect to electromagnetic forces that its mass plays with respect to gravity. The differences are, as already mentioned, that (i) unlike mass which is always positive, electric charge could be positive, negative or even zero in the neutral case; (ii) the mass appears in Newton's Second Law Motion, a role special to it and not shared by the electric charge at all.

Like mass in case of force of gravity, the electric charge of a body determines both how much of an electromagnetic force it causes on another charged body, and how much force it feels due to other charged bodies around it - the active and the passive aspects of electric charge. It is convenient and traditional to study such forces in the following sequence as we have done: (i) at

first one understands the properties of *electrostatic* forces, i.e. forces between electrically charged bodies, which are at rest (or moving at slow speeds, slow with respect to the speed of light which has the rather large value of 3×10^8m/s!); (ii) next one studies electric charges in motion, or electric currents and the thermal, chemical and magnetic effects they produce; (iii) this is followed by *magnetostatics.* i.e. for example, magnetic forces produced and felt by permanent magnets and magnetizable materials (in reality as we have seen all magnetic effects are ultimately due to electric charges in motion); (iv) finally one goes beyond these static or time-independent situations to study electromagnetic forces in full generality. Thus one goes progressively from electrostatics and magnetostatics which can be treated as separate subjects independent of one another, to true electromagnetism, where both electric and magnetic fields are simulataneously involved, after one another, and may be varying with time. It is at this stage that one learns about Fraday's Law of Induction, and of electromagnetic waves.

This is the approach we have taken in this book. However we ought to mention that when you study this subject at a somewhat higher level, for instance in college, it is much better to see at the very beginning the complete statement of the laws of electromagnetism, which are called the system of Maxwell's Equations, for the most general situations involving simulaneous electric and magnetic fields; one can then derive systematically the laws of electristatics and magnetostatics as special situations or special cases of the general laws. This approach is to be preferred because then one is impressed right at the beginning by the underlying unity of electric and magnetic phenomena, and also becuase this unity has been a part of our knowledge for more than a hundred years already.

One last point of comparison between gravitation and electromagnetism seems worth making. It is that even at the level of the very qualitative discussion we have here presented, one is left with the feeling that the laws of electric and magnetic forces are quite a bit more complicated than Newton's "simple" universal law of gravitation. This is true indeed, but then one must remember that the complete and proper understanding of gravitation is the one contained in the general theory of relativity, amd this is much more intricate than the laws of electricity and magnetism! Newton's law of gravity is just the form that Einstein's law reduces to in some limiting simple and situations.

Appendix D

UNITS IN ELECTRICITY AND MAGNETISM

You already know about the SI units from Chapter 2 of Class XI text book. The seven base SI units are metre (m), kilogram (kg), second(s), kelvin (K), ampére (A), candela (cd) and mole (mol). You are also familiar with several derived units in mechanics such as, for example, newton (N), joule (J), watt (W), pascal (Pa), poise etc. Though SI units have now been adopted internationally, other systems of units, e.g. the CGS system are also sometimes used for convenience. In mechanics, the basic formulas defining or relating various physical quantities are identical in SI and CGS system.Conversion from SI to CGS system is, therefore, a straightforward matter in mechanics.

Units in electricity and magnetism are somewhat more difficult to deal with for a number of reasons. First, to a beginning student, the formulas of electricity and magnetism contain a number of new, unfamiliar demensional constants such as ϵ_0, μ_0 and new physical quantities such as $\mathbf{E}, \mathbf{B}, L, C, R$ etc. Second, and more important, the basic quantities do nto look identical in different systems of units. Most work, especially in enegineering, is carried out in SI units but sometimes in certain other areas of work, systems of units other the SI are also employed for convenience or out of early practice. We shall restrict ourselves here only to the SI and another common system of units known as the Gaussian CGS system.

SI

In SI units current is considered a separate dimension, besides the familiar three dimensions of length, time and mass. (The remaining three SI units are irrelevant for the present discussion). The unit of current namely an ampére (A) is *defined* via the formula for mutual force per unit length between two infinitely long parallel current carrying conductors placed at a distance d apart

$$\frac{F}{\ell} = \propto \frac{I_1 I_2}{d}. \tag{1}$$

The constant of proportionality is written as $\mu_0/2\pi$ and given an *exact* numerical value $= 2 \times 10^{-7}$ i.e.

$$\frac{F}{\ell} = \frac{\mu_0}{2\pi}\,\frac{I_1 I_2}{d}. \tag{2}$$

This leads to the definition of the unit of current is SI:

1 ampere (A) *is that current which, if present in each of two parallel conductors of infinite length placed 1 m apart in vacuum, gives rise to a force on each conductor equal to* 2×10^{-7} N *per metre of length.*

With this definition,

$$\mu_0 \equiv 4\pi \times 10^{-7}\mathrm{NA}^{-2}. \tag{3}$$

The various SI units in electricity and magnetism are derived units in terms of m, kg, s and A. Let us consider them one by one.

1 coulomb (C) *is the charge flowing in 1s across a cross- section of a conductor carrying a constant current of 1* A

$$1\,\mathrm{C} = 1\,\mathrm{A\ s}. \tag{4}$$

Having fixed the unit of charge, the proportionality constant appearing in Coulomb's law is no longer a matter of choice but is determined through experiment

$$|\mathbf{F}| = \frac{k q_1 q_2}{r^2} \tag{5}$$

where k is found (not defined) to be equal to 8.987554×10^9 $\mathrm{Nm^2C^{-2}}$ or nearly 9.0×10^9 $\mathrm{Nm^2\ C^{-2}}$. k is usually written as $(1/4\pi\epsilon_0)$ so that Coulomb's law in SI units reads

$$|\mathbf{F}| = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2} \tag{6}$$

where $\epsilon_0 = 8.854185 \times 10^{-12}\ \mathrm{C^2\ N^{-1}m^{-2}}$.

Next, using the definition of electric field and potential

$$\mathbf{E} = \frac{\mathbf{F}}{q} \text{ and } V = \frac{W}{q} \tag{7}$$

the unit of electric field is fixed to be $1\mathrm{NC^{-1}}$ and the unit of electric potential is fixed to be

$$1\mathrm{V} = 1\ \mathrm{JC^{-1}}. \tag{8}$$

Clearly, the unit of electric field $1\ \mathrm{NC^{-1}}$ may also be written as $1\ \mathrm{Vm^{-1}}$.

Gauss's law in SI units reads

$$\oint \mathbf{E}\cdot\mathbf{ds} = \frac{q}{\epsilon_0} \tag{9}$$

where q is the charge enclosed by a closed surface S. You can easily verify that the dimensions of the two sides of Eq. (9) match.

How to define the unit of magnetic field? This is most easily done by considereing the magnetic force on a change q moving with velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$

$$\mathbf{F}_\mathrm{m} \propto q\mathbf{v}\times\mathbf{B}. \tag{10}$$

In SI units, the constant of proportionality in the above equation is *chosen* to be unity. The Lorentz force on a charge q in $\mathbf{E}$ and $\mathbf{B}$ is then given by

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v}\times\mathbf{B}. \tag{11}$$

This fixes the SI unit of magnetic field $\mathbf{B}$, called a tesla (T)

$$1\mathrm{T} = 1\mathrm{NA^{-1}m^{-1}}. \tag{12}$$

Equivalently, one can consider the formula for force on a current carrying conductor placed in a magnetic field and choose the constant of proportionality to be unity.

$$\mathbf{F} = I\boldsymbol{\ell}\times\mathbf{B} \tag{13}$$

where ℓ is in the direction of current. From Eq. (13), *1 tesla is that magnetic field in which a conductor of length 1* m *placed normal to the field and carrying a current of 1* A *experiences a force of 1* N.

Now the mutual force between two parallel current-carrying conductors arises due to the force experienced by one conductor in the presence of magnetic field produced by the other. Eqs (2) and (13) lead to the expression for magnetic field due to a long straight conductor at a point distant r from the conductor

$$|\mathbf{B}| = \frac{\mu_0 I}{2\pi r}. \tag{14}$$

The consants of proportionality in Eqs (2) and (13) having been chosen already, the over-all constant in Eq. (14) is forced on us and is not a matter of choice. A more general expression than Eq. (14) (Biot and Savart law) is given by

$$\mathbf{dB} = \frac{\mu_0}{4\pi}\frac{I\mathbf{d}\boldsymbol{\ell}\times\mathbf{r}}{r^3} \tag{15}$$

where $\mathbf{dB}$ is the magnetic field produced by a current element of length $\mathbf{d}\boldsymbol{\ell}$ at a point with position vector $\mathbf{r}$ from the element.

The magnetic flux Φ is the surface integral of $\mathbf{B}$.

$$\Phi = \oint_s \mathbf{B}\cdot\mathbf{ds} \tag{16}$$

The unit of Φ is 1 weber = 1 $\mathrm{Tm^2}$. (17)

Gauss's law of magnetism reads

$$\oint \mathbf{B}\cdot\mathbf{ds} = 0. \tag{18}$$

which is obviously consistent dimensionally, since zero can be assigned any dimension!

Having thus fixed the units and dimensions of $\mathbf{E}$ and $\mathbf{B}$ (and the dimensions of

ϵ_0, μ_0) the dimensions of constants appearing in Faraday's law of electromagnetic induction and in the term containing displacement current in the modified Ampére's law are fixed and not a matter of choice. The magnitude of these constants is fixed by experiment. In SI units, Faraday's law is

$$\mathcal{E} = \oint \mathbf{E} \cdot \mathbf{d}\ell = -\frac{d\Phi}{dt} \tag{19}$$

and the modified Ampére's law reads:

$$\oint \mathbf{B} \cdot \mathbf{d}\ell = \mu_0 I + \mu_0\epsilon_0 \frac{d}{dt} \int \mathbf{E.ds}. \tag{20}$$

It is worth noting that $1/\sqrt{\mu_0\epsilon_0}$ has the dimension and magnitude of the speed of light

$$c = \frac{1}{\sqrt{\mu_0\epsilon_0}}. \tag{21a}$$

With $\mu_0 = 4\pi \times 10^{-7}\mathrm{NA}^{-2}$, eq.(21a) gives

$$\epsilon_0 = \frac{10^7}{4\pi c^2}\mathrm{C^2N^{-1}m^{-2}}. \tag{21b}$$

[Eq.(21b) serves as a defining relation for ϵ_0 in the most recent conventions (not adopted in this book) wherein c is given an exact value $= 2.99792458 \times 10^8$ ms^{-1}].

Finally, the SI units of R, C and L namely 1 ohm, 1 farad and 1 henry respectively are fixed by their defining relations:

$$V = IR \text{ (Ohm's law) } 1\Omega = 1\mathrm{VA}^{-1} \tag{22}$$

$$\mathrm{Q = CV} \quad 1\mathrm{F} = 1\mathrm{CV}^{-1} \tag{23}$$

$$\mathcal{E} = -L\frac{di}{dt} \quad 1\mathrm{H} = 1\mathrm{VA^{-1}s} = 1\Omega\mathrm{s}. \tag{24}$$

The various derived units obtained above may be explicitly written in terms of the base units of SI by writing $1\mathrm{N} = 1$ kg m^2 s^{-2}, $1\mathrm{J} =$ 1 kg m^2 s^{-2}, $1\mathrm{C} = 1$ A s etc. For simplicity, we have restricted our discusion to equations of electromagnetism in free space only.

Gaussian CGS system

In this system, charge or current is not taken as an independent dimension. Rather, in Coulomb's law the proportionality constant is taken to be dimensionless and its magnitude is chosen to be unity.

$$|\mathbf{F}| = \frac{q_1 q_2}{r^2}. \tag{25}$$

The unit of charge called statcoulomb (statcoul) is thus a derived unit in terms of the three base units: cm, g,s

$$\begin{aligned} 1 \text{ statcoul} &= [1\mathrm{g \ cm \ s^{-2}cm^2}]^{1/2} \\ &= 1\mathrm{g^{1/2}cm^{3/2}s^{-1}}. \end{aligned} \tag{26}$$

Current is charge flowing per unit time. The unit of current is

$$1 \text{ statamp} = 1 \text{ statcouls}^{-1} \tag{27}$$

Using the same definitions as in Eq. (7), the unit of electric potential is

$$1\text{statvolt} = 1 \text{ erg statcoul}^{-1} \tag{28}$$

and the unit of electric field is 1 dyn statcoul^{-1} = 1 statvolt cm^{-1}.

Gauss's law in Gaussian CGS system, obtained from Eq. (25) and the definition of $\mathbf{E}$ reads

$$\oint_S \mathbf{E} \cdot \mathbf{ds} = 4\pi q. \tag{29}$$

The magnetic field $\mathbf{B}$ is again defined by the relation (Eq. 10)

$$\mathbf{F}_{\mathrm{m}} \propto q\mathbf{v} \times \mathbf{B}.$$

In the Gaussian CGS system, however, the constant of proportionality is chosen to be $1/c$, where c is nearly equal to 3.0×10^{10} cms^{-1}. The Lorentz force equation then reads

$$\mathbf{F} = q\mathbf{E} + \frac{q}{c}\mathbf{v} \times \mathbf{B}. \tag{30}$$

Clearly, **E** and **B** have the same dimensions in this system. However, for convenience, a different name is given to the unit of magnetic field, namely a gauss (G). Dimensionally

$$1\,\mathrm{G} = 1 \text{ statvolt cm}^{-1}. \tag{31}$$

The unit of magnetic flux Φ defined by Eq. (16) is 1 maxwell (Mx):

$$1\mathrm{Mx} = 1\mathrm{G\ cm}^2 \tag{32}$$

Having thus fixed the units of charge, **E**, **B**, etc., the proportionality constants in the remaining laws of electromagnetism are no longer a matter of choice; their magnitudes are fixed by experiment:

In the Gaussian CGS system, Biot and Savart law reads,

$$\mathbf{dB} = \frac{1}{c}\,\frac{I\mathbf{d}\boldsymbol{\ell}\times\mathbf{r}}{r^3} \tag{33}$$

which for a long straight conductor gives

$$|\mathbf{B}| = \frac{2I}{cr}. \tag{34}$$

Gauss's law of magnetism is, as before (Eq. 18)

$$\oint \mathbf{B}\cdot\mathbf{ds} = 0$$

The equation for force on a current carrying conductor in a magnetic field **B** is

$$\mathbf{F} = \frac{I}{c}\ell\times\mathbf{B} \tag{35}$$

where the proportionality constant $1/c$ follows from the choice of constant in the rsecond term of Eq. (30).

The force per unit length between two infinitely long parallel conductors is given by

$$\frac{F}{\ell} = \frac{2}{c^2}\,\frac{I_1I_2}{d} \tag{36}$$

which follows using Eq. (34) and (35).

In this system of units, Faraday's law of e.m. induction is

$$\mathcal{E} \equiv \oint \mathbf{E}\cdot \mathrm{d}\ell = -\frac{1}{c}\frac{d\Phi}{dt} \tag{37}$$

and the modified Ampere's law reads

$$\oint \mathbf{B}\cdot\mathbf{d}\boldsymbol{\ell} = \frac{4\pi}{c}I + \frac{1}{c}\frac{d}{dt}\int \mathbf{E}\cdot\mathbf{ds} \tag{38}$$

The defining relations for R, C and L are as before. Let us find out the units of these quantities in terms of the base units: cm, g, s

$$V = IR \quad .$$

The unit or resistance is 1 statvolt statamp^{-1}

$$\begin{aligned}
&= 1 \text{ erg statcoul}^{-1}\text{ statcoul}^{-1}\mathrm{s}\\
&= 1 \text{ gcm}^2\mathrm{s}^{-1}\ [\text{statcoul}]^{-1}\\
&= 1 \text{ gcm}^2\mathrm{s}^{-1}\cdot 1\mathrm{g}^{-1}\mathrm{cm}^{-3}\mathrm{s}^2\\
&= 1 \text{ cm}^{-1}\ \mathrm{s}^{-1}
\end{aligned} \tag{39}$$

Next

$$Q = CV.$$

The unit of capacitance is

$$\begin{aligned}
1 \text{ statcoul statvolt}^{-1} &= 1 \text{ statcoul}^2\text{erg}^{-1}\\
&= \text{gcm}^3\ \mathrm{s}^{-2}\cdot 1\ \mathrm{g}^{-1}\ \ \mathrm{cm}^{-2}\ \mathrm{s}^2\\
&= 1 \text{ cm.}
\end{aligned}$$

$$\mathcal{E} = -L\frac{di}{dt}. \tag{40}$$

The unit of inductance is

$$\begin{aligned}
\frac{(1\text{statvolt})\mathrm{s}}{1\text{statamp}} &= 1\mathrm{cm}^{-1}\mathrm{ss}\\
&= 1\mathrm{cm}^{-1}\mathrm{s}^2
\end{aligned} \tag{41}$$

Conversion from SI to Gaussian CGS system

The conversion factors for purely mechanical units are easy to obtain because the defining relations and dimensions of mechanical quantities are identical in the two systems. For example,

$$\begin{aligned}
1\mathrm{N} &= 1\mathrm{kg\ \ ms}^{-2} = 10^3\mathrm{g}10^2\mathrm{cms}^{-2}\\
&= 10^5\mathrm{g\ \ cms}^{-2} = 10^5\mathrm{dyn},\\
1\,\mathrm{J} &= 1\mathrm{Nm} = 10^5\mathrm{dyn}10^2\mathrm{cm}\\
&= 10^7\mathrm{dyn\ \ cm} = 10^7\mathrm{erg}.
\end{aligned}$$

For electrical and magnetic quantities, not only the units (i.e. their 'size') but also their dimensions differ from one system to another because in SI current is an independent dimension, whereas in Gaussian CGS system it is not. Obtaining conversion factors between the units of a given physical quantity in the two systems, therefore, requires some care. Let us start with Coulomb's law in SI

$$F = \frac{1}{4\pi\epsilon_0}\,\frac{q_1 q_2}{r^2}. \qquad \text{(SI)}$$

For $q_1 = q_2 = 1$ C; $r = 1$ m,

$$F = \frac{1}{4\pi\epsilon_0} = 9 \times 10^9 \text{N}.$$

Therefore, for $q_1 = q_2 = 1$ C,

$$\begin{aligned} r &= 1 \text{ cm } = 10^{-2}\text{m}, \\ F &= 9 \times 10^9 \times 10^4 = 9 \times 10^{-13}\text{N} \\ &= 9 \times 10^{18}\text{dyn}. \end{aligned} \qquad (42)$$

Now from Coulomb's law in Gaussian CGS system

$$F = \frac{q_1 q_2}{r^2} \qquad \text{Gaussian CGS,}$$

for $q_1 = q_2 = 1$ statcoul, $r = 1$ cm,

$$F = 1\text{dyn}. \qquad (43)$$

Comparison of Eq. (42) with Eq. (43) gives

$$(1\text{C})^2 = 9 \times 10^{18}(1 \text{ statcoul})^2$$

i.e.

$$1\,\text{C} = 3 \times 10^9 \text{ statamp}. \qquad (44)$$

This implies

$$\begin{aligned} 1\text{A} &= 1\text{C s}^{-1} = 3 \times 10^9 \text{ statcoul s}^{-1} \\ &= 3 \times 10^9 \text{statamp} \end{aligned} \qquad (45)$$

Next,

$$\begin{aligned} 1\text{V} &= 1\text{JC}^{-1} = \frac{10^7\text{erg}}{3 \times 10^9\text{statcoul}} \\ &= \frac{1}{300} \text{ statvolt}. \end{aligned} \qquad (46)$$

The conversion factor between the units of electric fields is similarly obtained

$$\begin{aligned} 1\text{Vm}^{-1} &= \frac{1 \text{ statvolt}}{300 \times 10^2 \text{ cm}} \\ &= \frac{1}{3} \times 10^{-4} \text{ statvolt cm}^{-1}. \end{aligned} \qquad (47)$$

To obtain conversion factor between the units of $\mathbf{B}$, consider the magnetic force part of the Lorentz force in SI:

$$\mathbf{F}_\text{m} = q\mathbf{v} \times \mathbf{B}. \qquad \text{(SI)}$$

For $q = 1$ C, $v = 1$ ms^{-1}, $B = 1$ T, with $\mathbf{v}$ normal to $\mathbf{B}$,

$$|\mathbf{F}_\text{m}| = 1\text{N}. \qquad (48)$$

Therefore, for

$$\begin{aligned} q &= 1 \text{ statcoul } = \frac{1}{3} \times 10^{-9}\text{C} \\ v &= 1\text{cms}^{-1} = 10^{-2}\text{ms}^{-1} \\ B &= 1\text{T with } \mathbf{v} \text{ normal to } \mathbf{B}, \\ |\mathbf{F}_\text{m}| &= \frac{1}{3}10^{-9} \times 10^{-2} \times 1 = \frac{1}{3} \times 10^{-11}\text{N} \\ &= \frac{1}{3} \times 10^{-6}\text{dyn}. \end{aligned} \qquad (49)$$

Now in the Gaussian CGS system,

$$\mathbf{F}_\text{m} = \frac{q}{c}\mathbf{v} \times \mathbf{B}, c = 3 \times 10^{10}\text{cms}^{-1}$$

so that for $q = 1$ statcoul, $v = 1$ cms^{-1}, $B = 1$ G, with $\mathbf{v}$ normal to $\mathbf{B}$,

$$|\mathbf{F}_\text{m}| = \frac{1}{3}10^{-10}\text{dyn}. \qquad (50)$$

Comparison of Eq. (49) with Eq.(50) gives

$$1\text{T} = 10^4\text{G}. \qquad (51)$$

It then follows that

$$\begin{aligned} 1\text{Wb} &= 1\text{Tm}^2 = 10^4\text{G} \times 10^4\text{cm}^2 \\ &= 10^8\text{Mx}. \end{aligned} \qquad (52)$$

Next,

$$\begin{aligned} 1\Omega &= \frac{1\text{V}}{1\text{A}} = \frac{(1/300) \text{ statvolt}}{3 \times 10^9 \text{ statamp}} \\ &= \frac{1}{9} \times 10^{-11} \text{ cm}^{-1}\text{s} \end{aligned} \qquad (53)$$

$$1\text{F} = \frac{1C}{1V} = \frac{3 \times 10^9 \text{ statcoul}}{1/300\text{statvolt}}$$

$$= 9 \times 10^{11}\text{cm} \qquad (54)$$

$$1\text{H} = 1\Omega\,\text{s} = \frac{1}{9} \times 10^{-11}\text{s cm}^{-1}\text{ s}$$

$$= \frac{1}{9} \times 10^{-11}\text{cm}^{-1}\text{s}^2. \qquad (55)$$

Prescriptions for conversion of equations of electricity and magnetism from SI to Gaussian CGS system can also be arrived at from above. We do not discuss them here. Besides the two systems of units considered in this appendix, a number of other systems of units (e.g. emu, esu etc.) are occasionally used in electricity and magnetism. The interested student may consult a more advanced text on this subject for learning these different systems of units.

ANSWERS TO EXERCISES AND ADDITIONAL EXERCISES

Chapter 7

7.3 1.6×10^{-3} V

7.4 7.5×10^{-6} V

7.5 100 V

7.7 (a) along abcd [motion causes increase in flux into the loop, so induced current tends to decrease the flux into the loop].

(b) along acba

(c) along adcb

(d) along abcd

Direction of induced current same as expected on magnetic force considerations. Note, induced current ceases when a loop is completely in or out.

7.8 (a) along pq

(b) along qp, along xy

(c) along xyz

(d) along zyx

(e) along xy

(f) No induced current since field lines lie in the plane of the loop.

7.10 1.5 V

7.11 Flux through each turn of the loop $= \pi r^2 B \cos \omega t$:

$$\mathcal{E} = -N\omega\pi r^2 B \sin\omega t;$$

$$\mathcal{E}_{\text{max}} = N\omega\pi r^2 B$$
$$= 20 \times 50 \times \pi \times 64 \times 10^{-4} \times 3.0 \times 10^{-2} = 0.603\text{V}$$

$\mathcal{E}_{av}$ is zero over a cycle.

$$I_{\text{max}} = 0.0603\text{A}$$

$$\text{Power loss} = (1/2)\mathcal{E}_{\text{max}} I_{\text{max}} = 0.036\text{W}$$

The induced current causes a torque opposing the rotation of the coil. An external agent (a rotor) must supply torque (and do work) to counter this torque in order to keep the coil rotating uniformly. Thus the source of the power dissipated as heat in the coil is the external rotor.

7.12 Induced emf $= (1/2)\omega B R^2$

$$= \frac{1}{2} \times 4\pi \times 0.4 \times 10^{-4} \times (0.5)^2$$
$$= 6.28 \times 10^{-5}\ \text{V}$$

The number of spokes is immaterial, because the emf's across the spokes are 'in parallel'

7.14 1.84 kV

7.15 400V

Answers to Additional Exercises

7.16 (a) along adcb (flux into increases during shape change, so induced current produces flux out).

(b) along adcb (flux out decreases during shape change, so induced current produces flux out).

7.17 (a) No; current is induced only if there is a *change* in the flux

linking the loop.

(b) No current is induced in *either* case. Current is induced due to changing magnetic (not electric) flux.

(c) In the rectangular loop. (For the circular loop, the rate of change of area of the loop inside the field region is not constant)

(d) a will be positive relative to b.

7.18 (i) 2.4×10^{-4} V lasting 2s

(ii) 0.6×10^{-4} V lasting 8s.

7.19 (a) Induced emf $= 8 \times 2 \times 10^{-4} \times 0.02 = 3.2 \times 10^{-5}$ V

Induced current $= 2 \times 10^{-5}$ A

Power loss $= 6.4 \times 10^{-10}$ W

Source of this power is the external agent responsible for changing the magnetic field with time.

7.20 Rate of change of flux due to explicit time variation in B

$= 144 \times 10^{-4}\,\text{m}^2 \times 10^{-3}\,\text{T s}^{-1}$

$= 1.44 \times 10^{-5}\,\text{Wb s}^{-1}$

Rate of change of flux due to motion of the loop in a non-uniform B

$= 144 \times 10^{-4}\,\text{m}^2 \times 10^{-3}\,\text{Tcm}^{-1}$
$\times 8\,\text{cm s}^{-1}$

$= 11.52 \times 10^{-5}\,\text{Wb s}^{-1}$

The two effects add up since both cause a *decrease* in flux along the positive z direction. Therefore, induced emf $= 12.96 \times 10^{-5}$V; induced current $= 2.88 \times 10^{-2} \simeq 2.9 \times 10^{-2}$ A. The direction of induced current is such as to *increase* the flux through the loop along +ve z-direction. If for the observer the loop moves to the right, the current will be seen to be anticlockwise. A proper proof of the procedure above is as follows.

$$\Phi(t) = \int_o^a a\, B(x,t)\mathrm{d}x;$$

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = a\int_o^a \mathrm{d}x \frac{\mathrm{d}B(x,t)}{\mathrm{d}t}$$

Using

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \frac{\partial B}{\partial t} + \frac{\partial B}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\partial B}{\partial t} + v\frac{\partial B}{\partial x}$$

We get

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = a\int_o^a \mathrm{d}x \left[\frac{\partial B}{\partial t}(x,t) + v\frac{\partial B}{\partial x}(x,t)\right] = A\left[\frac{\partial B}{\partial t} + v\frac{\partial B}{\partial x}\right],$$

where $A = a^2$.

The last step follows because $(\partial B/\partial t), (\partial B/\partial x), v$ are given to be constants in the problem. Even if you do not understand this formal proof (which require good familiarity with calculus), you will still appreciate that flux change can occur both due to motion of the loop as well as time - variations in magnetic field.

7.21 $Q = \int\limits_{t_i}^{t_f} I\mathrm{d}t$

$$= \frac{1}{R}\int_{t_i}^{t_f} \mathcal{E}\mathrm{d}t$$

$$= -\frac{N}{R}\int_{\Phi_i}^{\Phi_f} \mathrm{d}\Phi$$

$$= \frac{N}{R}(\Phi_i - \Phi_f)$$

for $N = 25, R = 0.50\Omega$,
$Q = 7.5 \times 10^{-3}$ C,
$\Phi_f = 0, A = 2.0 \times 10^{-4}\mathrm{m}^2$,
$\Phi_i = 1.5 \times 10^{-4}\mathrm{Wb}$;
$B = (\Phi_i/A) = 0.75\mathrm{T}$

7.22 (a) $|\mathcal{E}| = \mathrm{vBl} = 0.12\times0.50\times0.15 = 9.0$ mV; P positive end and Q negative end.

(b) Yes. When K is closed, the excess charge is maintained by the continuous flow of current.

(c) Magnetic force is cancelled by the electric force set up due to the excess charge of opposite signs at the ends of the rod.

(d) Retarding force $= IBl$

$$= \frac{9\mathrm{m}V}{9\mathrm{m}\Omega} \times 0.5\mathrm{T} \times 0.15 \text{ m}$$

$$= 75 \times 10^{-3}\,\mathrm{N}.$$

(e) Power expended by an external agent against the above retarding force to keep the rod moving uniformaly at 12 cm s^{-1} $= 75 \times 10^{-3} \times 12 \times 10^{-2} = 9.0 \times 10^{-3}$W. When K is open, no power is expended.

(f) $I^2R = 1 \times 1 \times 9 \times 10^{-3} = 9.0 \times 10^{-3}$ W. The source of this power is the power provided by the external agent as calculated above.

(g) Zero; motion of the rod does not cut across the field lines. [*Note*: length of PQ has been considered above to be equal to the spacing between the rails].

7.23 $B = \dfrac{\mu_0 NI}{\ell}$

(inside the solenoid away from the ends)

$$\Phi = \frac{\mu_0 NI}{\ell}A$$

Total flux linkage $= N\Phi$

$$= \frac{\mu_0 AN^2}{\ell}I$$

(ignoring end-variations in B).

$$|\mathcal{E}| = \frac{\mathrm{d}}{\mathrm{d}t}(N\Phi),$$

$$|\mathcal{E}|_{av} = \frac{\text{total change in flux}}{\text{total time}}$$

$$|\mathcal{E}|_{av} = \frac{4\pi \times 10^{-7} \times 25 \times 10^{-4}}{0.3 \times 10^{-3}} \times (500)^2 \times 2.5$$

$$= 6.5\,\mathrm{V}$$

7.24 $v = 1800\mathrm{km/h} = 500\mathrm{ms}^{-1}$,

vertical component of B

$= 5.0 \times 10^{-4} \sin 30°$

$= 2.5 \times 10^{-4}\,\mathrm{T}$

$\ell = 25\mathrm{m}$

$\mathcal{E} = 500 \times 2.5 \times 10^{-4} \times 25$

$= 3.1\,\mathrm{V}$

The direction of the wing is immaterial (as long as it is horizontal) for this answer.

Chapter 8

8.1 (a) 2.20 A

(b) 484 W

8.2 (a) $\frac{300}{\sqrt{2}} = 212.1\text{V}$

(b) $10\sqrt{2} = 14.1$ A

8.3 $1.28\ \pi\text{mH} = 4.02$ mH

8.4 $2.5\ \pi\text{mH} = 7.85$ mH

8.5 2.0×10^{-3}s

8.6 5.0×10^{-4}s

8.7 15.9 A

8.8 2.49 A.

8.9 Zero in each case

8.11 $125\ \text{s}^{-1}$; 25

8.12 Choke coil reduces voltage across the tube without wasting power. A resistor would waste power as heat.

8.15 $1.1 \times 10^{3}\ \text{s}^{-1}$

8.16 0.6J; same at later times

8.17 2000 W

8.18 $\nu = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}$ i.e. $C = \frac{1}{4\pi^2\nu^2 L}$

For L $= 200\mu\text{H}, \nu = 1200\text{kHz},$

$C = 87.9$ pF

For L $= 200\mu\text{H}, \nu = 800\text{kHz},$

$C = 197.8$ pF;

The variable condenser should have a range of about 88pF to 198 pF.

8.19 (a) Resonant frequency

$$\begin{aligned}\omega_0 &= (1/\sqrt{LC}) \\ &= (1/\sqrt{5 \times 80 \times 10^{-6}}) \\ &= 50 \text{ rad s}^{-1}\end{aligned}$$

(b) At $\omega = \omega_0$

$$\begin{aligned}|Z| &= \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \\ &= R = 40\Omega \\ I_0 &= \frac{V_0}{|Z|} = \frac{V_0}{R}\end{aligned}$$

(at resonance)

$$\begin{aligned}&= \frac{\sqrt{2} \times 230}{40} \\ &= 8.1\text{A}\end{aligned}$$

(c) Across L,

$$\begin{aligned}V_{Lrms} &= I_{rms} \times \omega_0 L \\ &= \frac{230}{40} \times 50 \times 5 \\ &= 1437.5\text{V}\end{aligned}$$

Across C

$$\begin{aligned}V_{Crms} &= I_{rms} \times \frac{1}{\omega_0 C} \\ &= \frac{230}{40} \times \frac{1}{50 \times 80 \times 10^{-6}} \\ &= 1437.5\text{V}\end{aligned}$$

Across L and C, total potential drop

$$V_{LCrms} = I_{rms}\left(\omega_0 L - \frac{1}{\omega_0 C}\right) = 0$$

Across R, $V_{Rrms} = 230\text{V}$

Answers to Additional Exercises

8.20 This exercise helps you appreciate the importance of the symmetry of mutual inductance : $M_{12} = M_{21}$ defined by $N_2\Phi_2 = M_{12}I_1$ and $N_1\Phi_1 = M_{21}I_2$. Suppose 1 represents the long solenoid and 2 the short solenoid. We are given I_2 and asked to find $N_1\Phi_1$. Direct calculation of this (i.e. M_{12}) will be a difficult task, because the short solenoid produces a complicated field whose flux through the long solenoid will be hard to calculate. Instead one uses $M_{12} = M_{21}$ and to calculate M_{12} , we need to find $N_2\Phi_2$ given I_1. This is easy since the long solenoid produces a (simple) uniform field inside given by $\mu_0 N_1 I_1/l_1$. Flux through each turn of the short solenoid is equal to $(\mu_0 N_1 I_1/\ell_1)\,\pi R_2^2$ where R_2 is the radius of the short solenoid.

Therefore

$$\frac{\mu_0 N_1 I_1}{l_1} \times \pi R_2^2 \times N_2 = M_{12} I_1$$

which gives

$$M_{21} = M_{12} = \frac{\mu_0 \pi R_2^2 N_1 N_2}{\ell_1}$$

The total flux linked with the long solenoid is

$$N_1\Phi_1 = M_{21} I_2 = \frac{\mu_0 \pi R_2^2 N_1 N_2}{\ell_1} I_2$$

Using the given data,

$$M_{21} = M_{12} = 2.96 \times 10^{-4}\text{H}$$

and

$$N_1\Phi_1 = 8.9 \times 10^{-4}\text{Wb}$$

8.21 (a) $B = \frac{\mu_0 N I}{2\pi r}$,

$$\Phi = \frac{\mu_0 N I}{2\pi r} A,$$

$$N\Phi = \frac{\mu_0 N^2 A}{2\pi r} I$$

$$\begin{aligned} L &= \frac{\mu_0 N^2 A}{2\pi r} \\ &= 4\pi \times 10^{-7} \\ &\times \frac{12 \times 10^{-4} \times 1200^2}{2\pi \times 0.15} \\ &= 2.3\text{mH} \end{aligned}$$

(b) $|\mathcal{E}| = \frac{\mathrm{d}}{\mathrm{d}t}(N_2\Phi_2)N_2$

$$\begin{aligned} &= \frac{\mu_0 N_1 N_2 A}{2\pi r}\frac{\mathrm{d}I_1}{\mathrm{d}t} \\ &= 4\pi \times 10^{-7} \times 300 \\ &\times \frac{1200 \times 12 \times 10^{-4} \times 2}{2\pi \times 0.15 \times 0.05} \\ &= 0.023\text{V}. \end{aligned}$$

8.22 (a) $\mathcal{E} = -L_{eq}\frac{\mathrm{d}I}{\mathrm{d}t}$,

$$\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2;$$

$$\mathcal{E}_1 = -L_1\frac{\mathrm{d}I}{\mathrm{d}t} - M\frac{\mathrm{d}I}{\mathrm{d}t}$$

$$\mathcal{E}_2 = -L_2\frac{\mathrm{d}I}{\mathrm{d}t} - M\frac{\mathrm{d}I}{\mathrm{d}t}$$

These equations imply $L_{eq} = L_1 + L_2 + 2M$. Here we have taken the currents to flow in the same sense in the two coils. If the series connection is such that the current flows in the opposite senses in two coils, $L_{eq} = L_1 + L_2 - 2M$.

(b) For the parallel connection, total current $I = I_1 + I_2$.

$$\mathcal{E} = -L_{eq}\frac{\mathrm{d}I}{\mathrm{d}t}$$

$$= -L_{eq}\left[\frac{dI_1}{dt}+\frac{dI_2}{dt}\right]$$

Also

$$\mathcal{E} = -L_1\frac{dI_1}{dt}$$
$$= -L_2\frac{dI_2}{dt}$$
(since $M = 0$)

These equations give

$$\frac{1}{L_{eq}} = \frac{1}{L_1}+\frac{1}{L_2}$$

8.23 (a) Energy expended by the source to build current from i to $i + di$ in time dt in an inductor $= (L\mathrm{d}i/\mathrm{d}t) \times i\mathrm{d}t = Li\mathrm{d}i$. Energy required to build current from 0 to I

$$= \int_0^I Li\,di$$
$$= \frac{1}{2}L\int_0^I d(i^2) = \frac{1}{2}LI^2.$$

(b) The energy stored in the two inductors is independent of the manner of building up currents in the coils. Let $I_2 = 0$ initially and let the current be built up from 0 to I_1 in coil 1. Energy required $= (1/2)L_1I_1^2$ as seen in (a). Now build up current in coil 2. Let the current increase from i_2 to $i_2 + di_2$ in time dt. Work done for coil $2 = i_2L_2\,(\mathrm{d}i_2/\mathrm{d}t)\mathrm{d}t$. The change in i_2 causes a flux change in 1 and induces emf in $1 = M(\mathrm{d}i_2/\mathrm{d}t)\mathrm{d}t$. Work done in d$t$ to maintain I_1 in 1 $= I_1M(\mathrm{d}i_2/\mathrm{d}t)$ dt. Thus total work done in raising I_2 from 0 to I_2 and maintaining I_1 in 1 is given by

$$L_2\int_0^{I_2} i_2di_2 + MI_1\int_0^{I_2} di_2$$
$$= \frac{1}{2}L_2I_2^2 + MI_1I_2$$

Hence the total energy stored in a pair of coupled coil is

$$\frac{1}{2}L_1I_1^2 + \frac{1}{2}L_2I_2^2 + MI_1I_2$$

Completing the square in I_1, this expression becomes

$$\frac{1}{2}L_1\left(I_1^2+\frac{2M}{L_1}I_1I_2+\frac{M^2}{L_1^2}I_2^2\right)$$
$$+\frac{1}{2}L_2I_2^2-\frac{1}{2}\frac{M^2}{L_1}I_2^2$$
$$= \frac{1}{2}L_1\left(I_1+\frac{M}{L_1}I_2\right)^2$$
$$+\frac{1}{2}\left(L_2-\frac{M^2}{L_1}\right)I_2^2$$

In order that the energy be non-negative for all values of I_1 and I_2 (including those values for which the first term is zero), a necessary and sufficient condition is that $L_2 > M^2/L_1$ i.e. $M^2 < L_1L_2$. More detailed discussion of this inequality can be found in Feynman Lectures (Vol.II) listed in the Bibliography.

8.24 As in 8.20, we should exploit here the symmetry $M_{12} = M_{21}$. Direct calculation of flux linking the bigger loop due to the field by the smaller loop will be difficult to handle. Instead, let us calculate the flux through the smaller loop due to a current I in the bigger loop. The smaller loop is so small in area that one can take the simple formula for

field B on the axis of the bigger loop and multiply B by the small area of the loop to calculate flux without much error. Let 1 refer to the bigger loop and 2 the smaller loop. Field B_2 at 2 due to I_1 in 1 is

$$B_2 = \frac{\mu_0 I_1 r_1^2}{2(x^2 + r_1^2)^{3/2}}$$

x: distance between the centres.

$$\Phi_2 = B_2 \pi r_2^2 = \frac{\pi \mu_0 r_1^2 r_2^2}{2(x^2 + r_1^2)^{3/2}} I_1$$

$$M_{21} = \frac{\pi \mu_0 r_1^2 r_2^2}{2(x^2 + r_1^2)^{3/2}} = M_{12}$$

$$\Phi_2 = M_{12} I_2 = \frac{\pi \mu_0 r_1^2 r_2^2}{2(x^2 + r_1^2)^{3/2}} I_2$$

Using the given data

$$\begin{aligned} M_{21} &= M_{12} = 4.55 \times 10^{-11}\text{H} \\ \Phi_1 &= 9.1 \times 10^{-11} \text{ Wb} \end{aligned}$$

8.25 (a) (i) The bulb in the R arm lights up earlier. Time constant is given by the ratio of self-inductance to resistance. This is negligible for the R arm. (ii) After the steady state, the self-inductance plays no role. The two arms of the circuit then behave identically because they have the same resistance and common applied voltage. They consume the same power. The bulbs will be equally bright in the steady state.

(b) The iron bar is a magnet. As the magnet falls down through the hollow region of the shell, the changing magnetic flux causes eddy currents in the shell which (according to Lenz's law) oppose the motion of the magnet.

(c) The self-inductance will be small due to the cancellation of induced emf effects. This is a special example of the situation in 8.22(a) when the winding is such that

$$\begin{aligned} L_{eq} &= L_1 + L_2 - 2M \\ &\simeq L + L - 2L \simeq 0 \end{aligned}$$

(d) Using Lenz's law it is easy to see that the electron speeds up if B increases with time. (This is, of course, expected physically since energy spent to increase B, say by increasing current in the electromagnet that produces B, should result in an increase in electrons's energy). The electron will not stay in the same circle in general because $R = (mv/eB)$ and v and B may not increase in proportion. However, it is possible to choose an appropriate non-uniform B and control its rate of increase so that R is fixed and unchanging with time. The machine called Betatron for accelerating charged particles is based on this principle.

(e) When the current in the coil of a large electromagnet is suddenly switched off, flux changes from a large value to zero in a very short time. Consequently, high voltages are induced across the open switch causing sparks and damaging the insulation. A small resistor in parallel provides a conducting path to the induced voltage thus avoiding sparks and other risks of high voltages.

8.26 $\frac{1}{C} = \frac{1}{4} + \frac{1}{6}$ i.e. $C = 2.4\mu\text{F}$

Time constant of the R-C circuit = $CR = 2.4 \times 10^{-6} \times 10^4 = 24\text{ms}$

During charging,

$q = CV(1 - e^{-t/RC})$

For $t = 10\text{s}$, $RC = 24\text{ms}$, $(t/RC) >> 1$ i.e. q is nearly CV i.e. the capacitor is fully charged to $Q_0 = CV = 2.4 \times 10^{-6} \times 18 = 43.2\mu\text{C}$.

During discharge $q = Q_0 e^{-t/RC}$. For $t = 48$ ms, $q = (43.2/e^2)\mu$ C,

$$V_1 = \frac{q}{C_1} = \frac{10.8}{e^2}\text{ V},$$

$$V_2 = \frac{7.2}{e^2}\text{ V}$$

$$V_1 = 1.46\text{V}, \quad V_2 = 0.975\text{ V}$$

Rounding off to 2 significant figures

$V_1 = 1.5\text{ V}, V_2 = 0.98\text{ V}$

8.27 (a) Yes, each equal to $(V/R) = (10/40) = 0.25$ A

(b) Current grows according to the relation:

$$I(t) = \frac{V}{R}(1 - e^{-(R/L)t})$$

At $t = (L/R)$,

$$I(t) = \frac{V}{R}(1 - e^{-1}) = \frac{V}{R}\left[1 - \frac{1}{e}\right]$$

Thus $(t_A/t_B) = (L_A/L_B)$ (since R is given to be the same) $= (10/0.05) = 200$

(c) Inductor A; recall energy required $= LI^2/2$

(d) Yes, power dissipated for each $= V^2/R$.

8.28 (a) Initial total energy

$$= \frac{Q_0^2}{2C} = \frac{10^{-4}}{2 \times 50 \times 10^{-6}} = 1.0\text{J}.$$

Yes, sum of the energies stored in L and C is conserved if R = 0.

(b) $$\omega = \frac{1}{\sqrt{LC}} = 10^3\text{rads}^{-1},$$
$$\nu = \frac{\omega}{2\pi} = 159\text{Hz}$$

(c) $q = Q_0 \cos \omega t$

(i) energy stored is completely electrical at

$$t = 0, \frac{T}{2}, T, \frac{3T}{2}$$

(ii) electrical energy is zero (i.e. energy stored is completely magnetic) at

$$t = \frac{T}{4}, \frac{3T}{4}, \frac{5T}{4}$$

where

$$T = \frac{1}{\nu} = 6.3\text{ms}$$

(d) At $t = T/8, 3T/8, 5T/8$, because

$$q = Q_0 \cos \frac{\omega T}{8} = Q_0 \cos \frac{\pi}{4} = \frac{Q_0}{\sqrt{2}}$$

Therefore electrical energy

$$= \frac{q^2}{2C} = \frac{1}{2}\left(\frac{Q_0^2}{2C}\right)$$

which is half the total energy.

(e) R damps out the LC oscillations eventually. The whole of the initial energy = 1.0J is eventually dissipated as heat.

8.29 For an LR circuit, if $V = V_0 \cos\omega t$

$$I = \frac{V_0}{\sqrt{R^2 + \omega^2 L^2}} \cos(\omega t - \delta),$$

where $\tan\delta = (\omega L/R)$

(a) $I_0 = \dfrac{V_0}{\sqrt{R^2+\omega^2L^2}}$

$= \sqrt{2} \times 240 \times$

$$\frac{1}{\sqrt{10^4 + (0.5)^2 \times 4\pi^2 \times 2500}}$$

$= 1.82\,\text{A}$

(b) V is maximum at $t = 0$, I is maximum at $t = (\delta/\omega)$. If δ is positive, this means current maximum lags behind voltage maximum by a time $= (\delta/\omega)$.

Now

$$\tan\delta = \frac{2\pi \times 50 \times 0.5}{100} = 1.571$$

$$\text{i.e. } \delta = \tan^{-1}(1.571) \simeq 57.5^\circ$$

$$\text{time lag} = \frac{57.5\pi}{180 \times 2\pi \times 50} = 3.2\text{ms}$$

8.30 For the high frequency $\omega = 2\pi \times 10^4$ rad s^{-1}

$$I_0 = \frac{\sqrt{2} \times 240}{\sqrt{10^4 + (0.5)^2 \times 4\pi^2 \times 10^8}} = 1.1 \times 10^{-2}\text{A}$$

(Note, R term negligible in the denominator above)

$$\tan\delta = \frac{2\pi \times 10^4 \times 0.5}{100} = 100\pi,$$

δ is close to $(\pi/2)$

I_0 is much smaller than for the low frequency case (8.29) showing thereby that at high frequencies L nearly amounts to an open circuit. In a dc circuit (after steady state) $\omega = 0$, so here L acts like a pure conductor.

8.31 For a C-R circuit, If $V = V_0 \cos\omega t$

$$I = \frac{V_0}{\sqrt{R^2 + \frac{1}{\omega^2 C^2}}} \cos(\omega t - \delta),$$

$$\tan\delta = -\frac{1}{\omega CR}$$

(a) $I_0 =$

$$\frac{\sqrt{2} \times 110}{\sqrt{1600 + \frac{1}{4\pi^2 \times 3600 \times 10^{-8}}}}$$

$= 3.23\text{A}$

(b) $\tan\delta =$

$$-\frac{1}{2\pi \times 60 \times 10^{-4} \times 40} = -0.6631$$

$$\delta = -33.5^\circ$$

$$\text{time lag} = \frac{\delta}{\omega} = -\frac{33.5\pi}{180 \times 2\pi \times 60} = 1.55\text{ms}$$

Negative sign means that current leads (i.e. voltage lags)

8.32 (a) $I_0 =$

$$\frac{\sqrt{2} \times 110}{\sqrt{1600 + \frac{1}{4\pi^2 \times 144 \times 10^6 \times 10^{-8}}}}$$

$= 3.88\text{A}$

(Note, C term negligible at high frequency)

(b) $\tan\delta =$

$$-\frac{1}{2\pi \times 12 \times 10^3 \times 10^{-4} \times 40}$$
$$= -\frac{1}{96\pi} \simeq -0.2$$

δ is nearly zero at high frequency.

We see that at high frequency C acts like a conductor. For a dc circuit after steady state, $\omega = 0$ and C amounts to an open circuit.

8.33 Effective impendance of the parallel LCR is given by

$$\frac{1}{Z} = \sqrt{\frac{1}{R^2} + (\omega C - \frac{1}{\omega L})^2}$$

which is minimum at

$$\omega = \omega_0 = \frac{1}{\sqrt{LC}}$$

Therefore $|Z|$ is maximum at $\omega = \omega_0$, and the total current amplitude is minimum.

In R branch,

$$I_{R\,\text{rms}} = \frac{V_{rms}}{R} = \frac{230}{40} = 5.75\text{A}$$

In L branch

$$I_{L\,\text{rms}} = \frac{V_{rms}}{\omega_0 L} = \frac{230}{50 \times 5} = 0.92\text{A}$$

In C branch,

$$I_{C\,\text{rms}} = V_{rms} \times \omega_0 C$$

$$= 230 \times 50 \times 80 \times 10^{-6}$$
$$= 0.92\text{A}$$

Note, total current $I_{rms} = 5.75$A, since the currents in L and C branch are 180° out of phase and add upto zero at every instant of the cycle.

8.34 (a) For $V = v_0 \cos\omega t$

$$I = \frac{V_0}{\left|\omega L - \frac{1}{\omega C}\right|} \times \cos\left(\omega t \pm \frac{\pi}{2}\right)$$

if $R = 0$

where $-$ve sign appears if $\omega L > 1/\omega C$ and +ve sign appears if $\omega L < 1/\omega C$.

For $\omega = 2\pi \times 50\text{rad s}^{-1}$, $V_0 = \sqrt{2} \times 230$ V,

$L = 80$mH, $C = 60\,\mu$F, this gives

$I_0 = 11.6A$, $I_{rms} = 8.24$ A

(b) $V^L_{rms} = I_{rms} \times \omega L = 207$ V

$V^C_{rms} = I_{rms} \times (1/\omega C) = 437$ V

(Note : 437 -207 = 230V (applied rms voltage). The voltages across L and C get subtracted because they are 180° out of phase.)

(c) Whatever be the current I in L, actual voltage leads current by $\pi/2$. Therefore average power consumed by $L = 0$.

(d) For C, voltage lags by $\pi/2$. Again average power consumed by $C = 0$

(e) Total average power absorbed $= 0$.

8.35 For $V = V_0 \cos\omega t$,

$$I = \frac{V_0}{\left[R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2\right]^{1/2}}$$

$\times \cos(\omega t - \delta)$

$$I_{rms} = \frac{230}{\sqrt{225 + (27.91)^2}}$$
$$= \frac{230}{31.68} = 7.26\text{A}$$

Average power to L = Average power to $C = 0$

Average power to R $= I_{\text{rms}}^2 R =$ 791W

Total power absorbed = 791W

8.36 (a) $I = I_0 \cos(\omega t - \delta)$, where

$$I_0 = \frac{V_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$$

$$\tan\delta = \frac{\omega L - \frac{1}{\omega C}}{R}$$

I_0 is maximum at $\omega = \omega_0$

$$\omega_0 = \frac{1}{\sqrt{LC}}$$
$$= \frac{1}{\sqrt{0.12 \times 480 \times 10^{-9}}}$$
$$= 4167\text{rad s}^{-1}$$
$$\nu_0 = \frac{\omega_0}{2\pi} = 663\text{Hz};$$
$$I_0^{\text{max}} = \frac{V_0}{R} = \frac{\sqrt{2} \times 230}{23}$$
$$= 14.1\text{ A}$$

(b) $P_{\text{av}} = (1/2)I_0^2 R$ which is maximum at the same frequency (663 Hz) for which I_0 is maximum.

$$P_{\text{av}}^{\text{max}} = \frac{1}{2}(I_0^{\text{max}})^2 R$$
$$= (I_{rms}^{\text{max}})^2 R = 2300\text{ W}$$

(c) At $\omega = \omega_0 \pm (R/2L)$ [Approximation good if $(R/2L) << \omega_0$]

$$\Delta\omega = \frac{R}{2L} = \frac{23}{0.24} = 95.8\text{rad s}^{-1}$$

i.e.

$$\Delta\nu = \frac{\Delta\omega}{2\pi} = 15.2\text{ Hz}$$

Power absorbed is half the peak power at $\nu = 648$ Hz and 678 Hz.

At there frequencies, current amplitude is $(1/\sqrt{2})$ times I_0^{max} i.e. current amplitude (at half the peak power points) is 10 A.

(d) $Q = \frac{\omega_0 L}{R} = \frac{4167 \times 0.12}{23}$
$= 21.7$

8.37 $\omega_0 = \frac{1}{\sqrt{LC}} = 111\text{rad s}^{-1}$,

$Q = \frac{\omega_0 L}{R} = 45$

To double Q without changing ω_0, reduce R to 3.7 Ω.

8.38 (a) Yes, The same is *not* true for rms voltage, because voltages across different elements may not be in phase. See, for example, answer to 8.34.

(b) Given a current in series LC, voltage in L leads current by $\pi/2$ phase and voltage in C lags current by $\pi/2$ phase. Therefore voltages in L and C differ by a phase of 180°. Given an applied voltage across parallel LC, current in L lags voltage by $\pi/2$ phase and current in C leads the voltage by $\pi/2$ phase. Therefore currents in L and C differ by a phase of π.

(c) To supply a given power low power factor means a large current is needed. This causes

larger heat losses due to the factor I^2R.

(d) Power factor = (R/Z). Many ac machines have inductive reactance. A capacitance of appropriate value reduces the net reactance so that Z approaches R.

(e) The high induced voltage when the circuit is broken is used to charge the capacitor, thus avoiding sparks etc.

(f) For dc, impedance of L is negligible and of C very high (infinite) so the dc signal appears across C. For high ω ac, impedance of L is high and that of C low. So the ac signal appears across L.

(g) For steady-state dc, L has no effect, even if it is increased by an iron core. For ac, the lamp will shine dimly because of additional impedance of the choke. It will dim further when the iron core is inserted which increases the choke's impedance.

(h) For dc, capacitor is an open circuit. The lamp will not shine at all, even if C is reduced. For ac, the lamp will shine because C 'conducts' ac. Reducing C will increase impedance of C and the lamp will shine less brightly than before.

8.39 (a) Amplitude of induced emf

$= NAB\omega$

$= 50 \times 2.5 \times 0.3 \times 60 = 2.25\,\text{kV}$

Maximum current drawn =

$$\frac{2.25 \times 10^3\text{V}}{500\Omega} = 4.5\text{A}$$

(b) Flux magnitude is maximum when current is zero, and it is zero when current is maximum. (Maximum flux = $BAN = 37.5$ Wb)

(c) Yes; what is required is relative motion. An alternator used for example, for bicycle lighting uses rotating pole pieces. Even for big ac generators, it is often more convenient to rotate the pole pieces instead of the coil.

8.40 The rotation of the motor in the magnetic field induces a back emf $\mathcal{E}'$ which opposes (according to Lenz's law) the applied emf $\mathcal{E}$. Thus

$$I = \frac{\mathcal{E} - \mathcal{E}'}{R}$$

$$\text{i.e. } 5.0 = \frac{200 - \mathcal{E}'}{8.5}$$

$$\text{or } \mathcal{E}' = 157.5V$$

Now power input = $\mathcal{E}'I = 1000$W

Power output = $\mathcal{E}I - I^2R = \mathcal{E}'I$

= 787.5 W

Efficiency = 78.75%

8.41 (a) If the armature is prevented from rotating (due to jamming) there is no back emf $\mathcal{E}'$ then the current is

$$\frac{200}{8.5} = 23.5\text{A}$$

This exceeds the limit 20A. The windings of the armature will burn out.

(b) The emf generated will be equal to the back emf in 8.40 i.e. 157.5V.

8.42 400.

8.43 Hydroelectric power $= h\rho g \times A \times v = h\rho g\beta$ where $\beta = Av$ is the flow (volume of water flowing per second across a cross-section)

Electric power available

$= 0.6 \times 300 \times \times 10^3 \times 9.8 \times 100$

$= 176\,\text{MW}.$

8.44 Line resistance $= 30 \times 0.5 = 15\Omega$

rms current in the line =

$$\frac{800 \times 1000\text{W}}{4000\text{V}} = 200\text{A}$$

(a) Line power loss $= 200 \times 200 \times 15 = 600$ kW

(b) Power supply by the plant = $800 + 600 = 1400$ kW

(c) Voltage drop on the line = $200 \times 15 = 3000$ V

The step-up transformer at the plant is 440V - 7000V.

8.45. Current $= \dfrac{800 \times 1000\text{W}}{40,000\text{V}} = 20\text{A}$

(a) Line power loss $= 20 \times 20 \times 15 =$ 6 kW

(b) Power supply by the plant = 806 kW

(c) Voltage drop on the line $= 20 \times 15 = 300$ V.

The step-up transformer is 440V - 40,300V. It is clear that percentage power loss is greatly reduced by high voltage transmission. In 8.44, this power loss is $(600/1400) \times 100 =$ 43%. In 8.45 it is only $(6/806) \times 100 = 0.74\%$.

Chapter 9

9.1 The speed in vacuum is the same for all: $c = 3 \times 10^8 \text{m s}^{-1}$

9.2 **E** and **B** lies in $x - y$ plane and are mutually perpendicular; 10 m

9.3 Wavelength band: 40m – 25m

9.5 10^9 Hz

9.7 Area covered

$= \pi \times 2R_E h$

$= \pi \times 2 \times 6.37 \times 10^6 \times 100 \text{ m}^2$

Polpulation covered

$= 12.74\pi \times 10^8 \times 10^{-3}$

$= 40$ lakhs

Answers to Additional Exercises

9.8 (a) $C = \epsilon_0 \text{ A/d}$

$$= \frac{8.85 \times 10^{-12} \times \pi \times 144 \times 10^{-4}}{5 \times 10^{-3}}$$

$= 80.1 \text{ pF}$

(b) $\frac{dQ}{dt} = C\frac{dv}{dt}$ Therefore,

$$\frac{dv}{dt} = \frac{0.15}{80.1 \times 10^{-12}}$$

$= 1.87 \times 10^9 \text{V s}^{-1}$

$$I_D = \epsilon_0 \frac{d}{dt}\phi_E.$$

Now across the capacitor

$\phi_E = EA$,

ignoring end corrections. Therefore

$$I_D = \epsilon_0 A \frac{dE}{dt}.$$

Now,

$$E = \frac{Q}{\epsilon_0 A},$$

therefore

$$\frac{dE}{dt} = \frac{I}{\epsilon_0 A}$$

which implies $I_D = I = 0.15 \text{ A}$.

(c) Yes, provided by 'current' we mean the sum of conduction and displacement currents.

9.9 (a) Consider a circle of radius r between the plates and co-axial with them (i.e., its centre lies on the axis of the plates and its plane is normal to the axis). By symmetry, **B** is tangential to the circle at every point and equal in magnitude over the circle. Therefore,

$$\oint \mathbf{B} \cdot d\vec{\ell} = 2\pi r B,$$

Using Ampére's law,

$2\pi r B = \mu_0 \times$

(current passing through the area enclosed by the circle).

$$= \mu_0 I_D \frac{r^2}{R^2} \quad r \leq R.$$

$$= \mu_0 I_D \quad r \geq R.$$

$$\text{Thus B} = \frac{\mu_0 r}{2\pi R^2} I_D \quad r \leq R$$

$$= \frac{\mu_0 I_D}{2\pi r} \quad r \geq R$$

(i) $B = 0$ on the axis $(r = 0)$

(ii) For $r = 6.5$ cm, $B = 1.35 \times 10^{-7}$ T

(iii) For $r = 15$ cm, $B = 2 \times 10^{-7}$ T

(b) B is maximum at $r = R$. From either formula above, at $r = R$

$$B = \frac{\mu_0 I_D}{2\pi R} = 2.5 \times 10^{-7}\text{T}.$$

9.10 (a) By Biot-Savart Law, magnetic field outside a straight current carrying wire is given by

$$B = \frac{\mu_0 I}{2\pi r}$$

Now since $I = I_D$, this formula is the same as in exercise 9.9 for $r \geq R$. Therefore, for $r = 12$ cm and 15 cm, answers are the same as in 9.9. For $r = 6.5$ cm ($r < R$), the two formulas differ. Here $B = 4.6 \times 10^{-7}$ T, greater than in 9.9.

(b) B is maximum at the surface of the wire,

$$B = \frac{4\pi \times 10^{-7} \times 0.15}{2\pi \times 10^{-3}} = 3.0 \times 10^{-5}\text{ T}$$

which is much greater than the maximum value of B in 9.9.

(c) Yes; since $I = I_D$, magnetic field configurations are identical for $r \geq R$.

9.11 Due to leaking, there is flow of +ve charge from the +ve plate to the -ve plate (or flow of -ve charge in the reverse direction). Thus the conduction current within the plates is from the +ve plate to the -ve plate. Now, the displacement current is

$$I_D = \epsilon_0 A \frac{\mathrm{d}E}{\mathrm{d}t} = \epsilon_0 A \frac{1}{\epsilon_0 A}\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{\mathrm{d}Q}{\mathrm{d}t}$$

But note here $\mathrm{d}Q/\mathrm{d}t$, $\mathrm{d}E/\mathrm{d}t < 0$. That is, I_D is opposite to the direction of electric field. Thus I_D has the same magnitude ($= 1.5 \times 10^{-8}$ A) as the conduction current but opposite direction. The total 'current' therefore is identically zero. This is true for any cross-section between the plates. Using Ampéres law as in 9.9 (with I_D there replaced by $I = I_C + I_D = 0$) the magnetic field within the plates is zero at all points.

9.12 (a) $I_{\text{rms}} = V_{\text{rms}}\, \omega C$

$= 230 \times 300 \times 10^{-10}$

$= 6.9\mu$ A

(b) Yes; The derivation in exercise 9.8(b) is true even if I is oscillating in time.

(c) The Formula

$$B = \frac{\mu_0}{2\pi}\ \frac{r}{R^2}\ I_D$$

goes through even if I_D (and therefore B) oscillates in time. The formula shows they oscillate in phase. Since $I_D = I$, we have,

$$B_0 = \frac{\mu_0}{2\pi}\ \frac{r}{R^2}\ I_0$$

where B_0 and I_0 are the amplitudes of the oscillating magnetic field and current respectively. $I_0 = \sqrt{2} I_{\text{rms}} = 9.76\mu$A. For $r = 3$ cm, $R = 6$ cm, $B_0 = 1.63 \times 10^{-11}$ T.

9.13 (a) (1) Gauss's law

(2) No particular name associated with this law

(3) Faraday's law

(4) Ampére's law (the second term on the right hand side of

this equation is the 'displacement current' due to Maxwell).

(b) Eqs (1) and (4) contain the sources q, I; Eqs. (2) and (3) do not. To obtain equations in source free region, simple put $q = I = 0$.

(c) Put the right hand side of Eq. (3) and the second term on the right hand side of Eq. (4) equal to zero, and take all other quantities time - independent.

(d) Eq(2) is based on the fact that monopoles do not exist. If they did, the RHS would contain a term, say q_m representing magnetic monopole strength, analogous to Gauss' law. Further, Eq. (3) would also be modified. An additional term $\mathbf{I}_m$ representing the current due to flow of magnetic charge would have to be included on the right hand side of Eq. (3), analogous to the electric charge current of Eq. (4). All this is, of course,, based on the expectation of symmetry of form of the equations for **E** and **B**. Nature may never show up monopoles, or else even if monopoles exist, the actual modifications of Maxwell's equations might be very different.. Who knows?

(e) Eq. (2)

(f) Eq. (3); if **E** is time-independent, $\oint \mathbf{E}.\mathbf{dl} = 0$.

(g) Yes, we can. Maxwell's equations can be cast as differential equations valid at every point in space and at every instant. You will learn them in more advanced courses.

(h) Maxwell's equations are true universally in all media (see, however, answer to (i) below). But in macroscopic media, it is usually convenient to write down equations for averages of **E** and **B** over regions which are small macroscopically but large enough to contain a very large number of atoms etc. The resulting macroscopic Maxwell's equations are of great practical use. You will learn them in your undergraduate courses.

(i) Maxwell's equations are basic laws of classical electromagnetism. They are true in all media and for any values of **E**, **B**, q, I etc. *within the domain of validity of classical electromagnetism.* The precise domain of validity is hard to specify and need not concern us here.

9.14 Let **E** be in the x-direction and **B** in the y-direction.

(a) Consider a rectangular loop in the $x-z$ plane with one side of length parallel to **E**. At the instant under consideration, the rectangle is partially on the left and partially on the right of the wavefront. Rate of change of magnetic flux $= Blc$. The line integral of **E** is El. Equating the two, $E = Bc$.

(b) Consider a similar ractangle in the $y-z$ plane. Rate of change of electric flux $= Elc$. The line integral of **B** is Bl. From

Ampére's law, $Bl = \mu_0\epsilon_o\ Elc$ i.e. $B = \mu_0\epsilon_0 cE$. Combine this with (a) to get $C = 1/\sqrt{\mu_0\epsilon_0}$.

9.15 (a) The speed of a wave is expected (by common sense) to be different for different observers in relative motion. This is so far the speed of a sound wave, for example. When we get a *constant* in the Maxwell's equations that is interpreted as the speed of electromagnetic waves, we are puzzled by the equation: which observer does the speed c refer to? There is no medium necessary for em wave propagation. So we are not able to think of a natural choice for the observer (say the one at rest with respect to the medium) relative to which the speed c is defined.

(b) The hypothesis of ether is natural attempt to deal with the above puzzling question. We say that the speed c refers to an observer at rest with respect to a medium (ether) that permeates all space and has very special properties. The ether hypothesis also allows us to think of electromagnetic waves as arising from changes in the properties of the medium – a picture that is less abstract than 'fields travelling in vacuum'.

(c) Maxwell's equations are *not* modified by special relativity. Rather, armed with special relativity, they are true for all (inertial) observers in uniform relative motion. The ether hypothesis is now known to be incorrect and c occuring in Maxwell's equations refers to the speed of electromagnetic waves in vacuum relative to any (inertial) observer. For processes which need quantization of the electromagnetic field (i.e. the photon picture of radiation). Maxwell's equations need to be handled in ways which are basically and non-trivially different from classical physics, and which cannot be discussed here.

9.16 (a) $\lambda = (c/\nu) = 1.5 \times 10^{-2}$m

(b) $B_0 = (E_0/c) = 1.6 \times 10^{-7}$ T

(c) Energy density in $\mathbf{E}$ field,

$$u_E = \frac{1}{2}\epsilon_0 \mathrm{E}^2.$$

Energy density in $\mathbf{B}$ field,

$$u_B = \frac{1}{2\mu_0}B^2$$

Using $E = cB$ and $c = 1/\sqrt{\mu_0\epsilon_0}$, $u_E = u_B$.

9.17 (a) There is no contradiction. Field lines *inside* the bar magnet go away from S and towards N. The net flux of $\mathbf{B}$ over any surface *fully* enclosing N or S must be identically zero.

(b) Not necessarily. A displacement current (such as between the plates of a capacitor that is being charged) can also produce loops of $\mathbf{B}$.

(c) Not Nècessarily. All that is needed is that the total *electric flux* through the area enclosed by the loop should vary in time. The flux change may arise from any *portion* of the area. Elsewhere **E** or d**E**/dt may be zero. In particular, there need be no electric field at the points which make the loop.

(d) We have seen in 9.9(b) how small the magnetic field is due to displacement current. The magnitude is too small to be easily observable. Of course, we can increase the effect by increasing the displacement current (in an a.c. circuit this can be done by increasing ω). The effect of induced electric field due to changing magnetic flux, on the other hand, can be increased simply by taking more and more number of turns in the coil. The induced emf's in different turns of the same coil add up in series.

(e) Increase in frequency causes decrease in impedence of the capacitor and consequent increase in the current which equals displacement current between the plates.

(f) The waves must satisfy a boundary condition. The electric field should be zero on the walls of the conductor. This restricts the modes possible. (It is something like the restricted modes of a string fixed at two ends).

9.18 Photon energy (for $\lambda = 1$ m)

$$= \frac{6.63 \times 10^{-34} \times 3 \times 10^{8}}{1.6 \times 10^{-19}} \text{ eV}$$
$$= 1.24 \times 10^{-6} \text{ eV}$$

Photon energy for other wavelengths in the figure for em spectrum can be obtained by multiplying appropriate powers of ten. Energy of a photon that a source produces indicates the spacings of the relevant energy levels of the source. For example, $\lambda = 10^{-12}$ m corresponds to photon energy $= 1.24 \times 10^{6}$ eV $= 1.24$ MeV. This indicates that nuclear energy levels (transition between which cause γ-ray emission) are typically spaced by 1 MeV or so. Similarly, a visible wavelength $\lambda = 5 \times 10^{-7}$ m corresponds to photon energy $= 2.5$ eV. This implies that energy levels (transitions between which give visible radiation) are typically spaced by a few eV.

9.19 A body at temperature T produces a continuous spectrum of wavelengths. For a black body, the wavelength corresponding to maximum intensity of radiation is given, according to Plank's law, by the relation: $\lambda_m = 0.29$ cm K$/T$. For $\lambda_m = 10^{-6}$ m, $T = 2900$ K. Temperatures for other wavelengths can be found. These numbers tell us the temperature ranges required for obtaining radiations in different parts of the em spectrum. Thus to obtain visible radiation, say $\lambda = 5 \times 10^{-7}$ m the source should have a temperature of about 6000 K. Note, a lower temperature will also produce this

wavelength but not with maximum intensity.

9.20 (i) Radio (short wavelength end)

(ii) Radio (short wavelength end)

(iii) Microwave

(iv) Visible (Yellow)

(v) X-rays (or soft γ–ray) region.

9.21 (a) Ionosphere reflects waves in these bands.

(b) Television signals are not properly reflected by the ionosphere (see text). Therefore, reflection is effected by satellites.

(c) Atmosphere absorbs X-rays, while visible and radiowaves can penetrate it.

(d) It absorbs ultraviolet radiations from the sun and prevents it from reaching the earth's surface and causing damage to life.

(e) The temperature of the earth would be lower because the Greenhouse effect of the atmosphere would be absent.

(f) The clouds produced by global nuclear war would perhaps cover substantial parts of the sky preventing solar light from reaching many parts of the globe. This would casue a 'winter'.

9.22 What you must realize here is that not all the quantities appearing in these questions have dimensions assigned to them from 'outside' (i.e. from independent equations and definitions). Some of the dimensions have to be fixed up from a few of the equations themselves, and the remaining equations (or terms in the equations) can then be checked for dimensional consistency. The dimension of ϵ_0 is given from Coulomb's law to be

$$[\epsilon_0] = \mathrm{C^2N^{-1}m^{-2}}$$

The Dimensions of E is fixed by its definition as force per unit charge

$$[\mathrm{E}] = \mathrm{N\ C^{-1}}$$

The dimensions of μ_0 may be fixed up using the first terms on the right hand side of Ampére's law (Eq. 4) and the dimension of current:

$$[I] = \mathrm{C\ s^{-1}}$$
$$[\mu_0] = \mathrm{NC^{-2}\,s^2}$$

Equipped with these dimensions, we can now check the dimensional consistency of Eq. (1), Eq. (3) and the second term of Eq. (4). The Lorentz force equation, and the first term of the right of Eq. (4) have been already used in assigning dimensions, so there is no meaning of checking their consistency. Eq. (2) is automatically consistent since 0 can be assigned any dimension!

Eq(1):

$$\left[\oint \mathbf{E}\cdot \mathbf{ds}\right] = \mathrm{N\ C^{-1}\ m^2}$$

$$\frac{q}{\epsilon_0} = \frac{\mathrm{C}}{\mathrm{C^2\ N^{-1}\ m^{-2}}} = \mathrm{N\ C^{-1}\ m^2}.$$

Eq.(3):

$$\left[\oint \mathbf{E.dl}\right] = \mathrm{N\ C^{-1}\ m}$$

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\int \mathbf{B}\cdot\mathbf{ds}\right] = \mathrm{N\,C^{-1}\ m^{-1}s}\ \frac{\mathrm{m^2}}{\mathrm{s}}$$

$$= \mathrm{N\,C^{-1}m.}$$

Eq.(4):

$$\left[\oint \mathbf{B.dl}\right] = \mathrm{N\,C^{-1}s.}$$

Second term on the right of Eq. (4)

$$\left[\mu_0\epsilon_0\frac{\mathrm{d}}{\mathrm{d}t}\left(\int \mathbf{E}\cdot\mathbf{ds}\right)\right]$$

$$= \mathrm{NC^{-2}s^2\ C^2N^{-1}m^{-2}\ NC^{-1}\ m^2\ s^{-1}}$$

$$= \mathrm{N\ C^{-1}\ s.}$$

Thus the above equations are dimensionally consistent.

Chapter 10

10.2 (a) Spherical

(b) plane

(c) plane (a small area on the surface of a large sphere is nearly planar)

10.4 5000 Å, 6×10^{14}Hz; 45°

10.6 (a) $2.0 \times 10^8 \text{ms}^{-1}$ (use $v = \frac{c}{n}$)

(b) No. The refractive index, and hence the speed of light in a medium, depends on wavelength. [When no particular wavelength or colour of light is specified, we may take the given refractive index to refer to yellow colour]. Now we know violet colour deviates more than red in a glass prism i.e. $n_v > n_R$. Therefore the violet component of white light travels slower than the red component.

10.7 400 nm, 5×10^{14} Hz, 2×10^8 ms^{-1}

10.9 Incoherent

10.10 $\lambda = \dfrac{1.2 \times 10^{-2} \times 0.28 \times 10^{-3}}{4 \times 1.4}\text{m}$

$= 600 \times 10^{-9}\text{m} = 600\text{nm}$

10.11 $\tan^{-1}(1.5) \simeq 56.3°$

10.12 (a) The factors by which the electric field gets multipled (after the first polaroid) are cos 45° and cos 45° for the second and third. Multiplying these and squaring, the fractional intensity transmitted after the first polaroid is 1/4.

10.13 13.2×10^{-2} rad $\simeq 7.56°$

10.14 40 m

10.15 12.5 cm

10.16 Use the formula

$$\lambda' - \lambda = \frac{v}{c}\lambda$$

$$\text{i.e. } v = \frac{c}{\lambda}(\lambda' - \lambda)$$

$$= \frac{3 \times 10^8 \times 15}{6563}$$

$$= 6.86 \times 10^5\text{ms}^{-1}$$

Answers to Additional Exercises

10.17 (a) Reflected light: (wavelength, frequency, speed same as incident light) $\lambda = 589$ nm, $\nu = 5.09 \times 10^{14}$ Hz, $c = 3.00 \times 10^8$ ms^{-1}.

(b) Refracted light: (frequency same as the incident *frequency*) $\nu = 5.09 \times 10^{14}$Hz $v = (c/n) = 2.26 \times 10^8\text{ms}^{-1}$ $\lambda = (v/\nu) = 444$ nm

10.18 In Newtons's corpuscular (particle) picture of refraction, particles of light incident from a rarer to a denser medium experience a force of attraction normal to the surface. This results in an increase in the normal component of the velocity but the component along the surface is unchanged. This means

$$c \sin i = v \sin r$$

or $\frac{v}{c} = \frac{\sin i}{\sin r} = n.$

Since $n > 1, v > c$.

The prediction is *opposite* to the experimental result: $(v < c)$. The wave picture of light is consistent with the experiment

10.19 With the point object at the centre, draw a circle touching the mirror. This is a plane section of the spherical wave front from the object that has just reached the mirror. Next draw the locations of this same wavefront after a time t in the presence of the mirror, and in the absence of the mirror. You will get two arcs symmetrically located on either side of the mirror. Using simple geometry, the centre of the reflected wavefront (the image of the object) is seen to be at the same distance from the mirror as the object.

10.20 (a) The speed of light in vacuum is a universal constant independent of all the factors listed and anything else. In particular, note the surprising fact that it is independent of the relative motion between the source and the observer. This fact is a basic axiom of Einstein's special theory of relativity.

(b) Dependence of the speed of light in a medium:

(i) does not depend on the nature of the source (wave speed is determined by the properties of the medium of propagation. This is also true for other waves e.g. sound waves, water waves etc).

(ii) independent of the direction of propagation for *isotropic media.*

(iii) independent of the motion of the source relative to the medium but depends on the motion of the observer relative to the medium.

(iv) depends on wavelength

(v) independent of intensity. [For high-intensity beams, however, the situation is more complicated and need not concern us here].

10.21 Sound waves require a medium for propagation. Thus even though the situations (i)and (ii) may correspond to the same relative motion (between the source and the observer), they are not identical physically since the motion of the observer relative to the medium is different in the two situations. Therefore we cannot expect Doppler formulas for sound to be identical for (i) and (ii). For light waves in vacuum, there is clearly nothing to distinguish between (i)and (ii). Here only the relative motion between the source and the observer counts and the relativistic Doppler formula is the same for (i) and (ii). For light propagation in a medium, once

again like for sound waves, the two situations are not identical and we should expect the Doppler formulas for this case to be different for the two situations (i) and (ii).

10.22 (a) Reflection and refraction (scattering in general) arise through interaction of incident light with the atomic constituents of matter. Atoms may be viewed as oscillators which take up the frequency of the external agency (light) causing the forced oscillations. The frequency of light emitted by a charged oscillator equals its frequency of oscillation. Thus the frequency of scattered light equals the frequency of incident light. The fact can also be explained more mathematically without using the atomic picture. At any interface between the two media, the electric (and magnetic) fields must satisfy certain boundary conditions. Frequency determines the time-dependence of fields. If incident, reflected and refracted frequencies were not equal, the same boundary conditions would not be satisfied for all times.

(b) No. Energy carried by a wave depends on the amplitude of the wave, not on the speed of wave propagation.

(c) A pulse can be viewed as being made of harmonic waves with a large range of wavelengths. Since the speed of propagation in a medium depends on wavelength, different wavelength components of the pulse travel with different speeds. The pulse will not retain its shape as it travels through the medium.

(d) For a given frequency, intensity of light in the photon picture is determined by the number of photons per unit area.

(e) The speed of light in water is not independent of the relative motion between the observer and the medium. We might expect the answer to be $(c/n) + v$. The correct answer according to special relativity (and experiments) is $(c/n) + v[1 - (1/n^2)]$ for $v << c$.

10.23 (a) Angular separation of the fringes remains constant ($= \lambda/d$). The actual separation of the fringes increases in proportion to the distance of the screen from the plane of the two slits.

(b) The separation of the fringes (and also angular separation) decreases. See, however, the condition mentioned in (d) below.

(c) The separation of the fringes (and also angular separation) decreases. See, however, the condition mentioned in (d) below.

(d) Let s be the size of the source and S its distance from the plane of the two slits. For interference fringes to be seen, the condition $s/S < \lambda/d$

should be satisfied; otherwise, interference patterns produced by different parts of the source overlap and no fringes are seen. Thus as S decreases (i.e. the source slit is brought closer) the interference pattern gets less and less sharp, and when the source is brought too close for this condtion to be valid, the fringes disappear. Till this happens, the fringe separation remains fixed.

(e) Same as in (d). As the source-slit width increases, fringe pattern gets less and less sharp. When the source slit is so wide that the condition $s/S \leq \lambda/d$ is not satisfied, the interference pattern disappears.

(f) The angular size of the central diffraction band due to each slit is about λ/S' where S' is the width of each of the two slits. S' should be sufficiently small so that these bands are wide enough to overlap and thus produce interference. This means $\lambda/S' >> \lambda/d$ i.e. the width of each slit should be considerably smaller than the separation between the slits. When the slits are so wide that this condition is not satisfied, fringes are not seen. However, increase in the width of the slits does improve the brightness of the fringes. Thus, in practice, the two slits should be wide enough to allow sufficient light to pass through but narrow enough to cause enough diffraction from each slit to enable wavefronts from the two slits to overlap and interfere.

(g) The interference patterns due to different component colours of while light overlap (incoherently). The central bright fringes for different colours are at the same position. Therefore, the central fringe is white. Since blue colour has the lower λ, the fringe closest on either side of the central white fringe is blue; the farthest is red. After a few, no clear fringe pattern is seen.

10.24 (a) Light waves coming directly from the source S and the reflected waves (which appear to come from the image S') interfere to produce a fringe pattern.

(b) To ensure that the separation between the two coherent sources S and S' is small, as required in a Young's double slit experiment.

(c) Reflection by mirror causes a phase change of 180^o which is equivalent to a change in path length by half a wavelength.

10.25 (a) Let the separation between the plates at a distance x from the joining line be y

$$y = \frac{x \times S}{\ell}$$

Now the condition for destructive interference (see (b)) is:

$$2y = n\lambda \quad \text{i.e.} \quad x = \frac{\ell}{2S} \times n\lambda$$

Therefore, the separation between the fringes is :

$$\Delta x = \frac{\ell\lambda}{2S}\Delta n = \frac{\ell\lambda}{2S}(\Delta n = 1)$$

(b) Reflection from the upper surface of the wedge causes no phase change (denser to rarer medium); reflection from the lower surface of the wedge causes a phase change of 180^o (rarer to denser medium). Hence the fringe along the line of contact is dark.

(c) Glass is denser than water. So the line of contact is still a dark fringe. But wavelength in water is less than that in air by a factor of 1.33. So the fringe separation is reduced by this factor.

(d) Choose the upper plate material, medium filling the wedge, and the lower plate material in increasing (or decreasing) order of refractive index.

10.26 (a) Straight lines parallel to the slits

(b) Straight lines parallel to the line of contact of the plates forming the air wedge.

(c) Straight lines parallel to the slits

(d) Here the two coherent sources are the images of the lamp (approximately a point source) in the front and back surfaces of the sheet. They produce circular fringes.

(e) Concentric circular fringes (Newton's rings) with the centre at the point of contact of the lens and the plate.

10.27 (a) The angular separation between the central bright band and the first dark band is λ/d. The angular separation between the two dark bands on either side of the central bright band is therefore $2\lambda/d$. Therefore, the actual separation between the two dark bands is $(2\lambda/d) \times D$. (D: distance of the screen). For the given data, this equals 4.68 mm.

(b) A circular hole produces circular diffraction fringes. The angular separation between the central bright band and the first dark band in this case is $1.22\ \lambda/d$ (Take this result without proof). The answer therefore modifies to $1.22 \times 4.68 =$ 5.71 mm

10.28 (a) The size reduces by half according to the relation: size $\sim \lambda/d$. Intensity increases four fold.

(b) The intensity of interference fringes in a double-slit arrangement is modulated by the diffraction pattern of each slit.

(c) Waves diffracted from the edge of the circular obstacle interfere constructively at the centre of the shadow producing a bright spot.

(d) For diffraction or bending of waves by obstacles/apertures by a large angle, the size of the latter should be comparable to

wavelength. If the size of the obstacle/aperture is much too large compared to wavelength, diffraction is by a small angle. Here the size is of the order of a few metres. The wavelength of light is about 5×10^{-7} m, while sound waves of say 1 kHz frequency have wavelength of about 0.3m. Thus sound waves can bend around the partition while light waves cannot.

(e) Justification based on what is explained in (d). Typical sizes of aperatures involved in ordinary optical instruments are much larger than the wavelength of light.

10.29 (a) Interference of the direct signal received by the antenna with the (weak) signal reflected by the passing aircraft.

(b) Light waves reflected from the upper and lower surfaces of a thin film interfere. Since the condition for constructive or destructive interference (bright or dark fringe) is wavelength dependent, coloured fringes are observed.

(c) Light wave reflected from the upper surface of the air film (denser to rarer medium) suffers a phase change of 180^o. Therefore the central fringe in the reflected light is dark. Transmitted light suffers no phase change at either surface. Hence the central fringe in the transmitted light is bright.

(d) If for a particular colour, interference is constructive in the reflected light, it is destructive in the transmitted light and vice versa. This is due to 180^0 change of phase in one of the reflected waves as mentioned in (c). Thus, if a particular colour produces a bright fringe in the reflected light, this colour will be reduced in the transmitted light from the same point. The coloured patterns in the reflected and the transmitted light, are therefore, complementary.

(e) Superposition principle follows from the linear character of the (differential) equation governing wave motion. If y_1 and y_2 are solutions of the wave equation, so is any linear combination of y_1 and y_2. When the amplitudes are large (e.g. high intensity laser beams) and non-linear effects are important, the situation is far more complicated and need not concern us here.

10.30 The electric field components in the two sets of axes are related by

$$E_x = E'_x \cos\theta - E'_y \sin\theta$$
$$E_y = E'_x \sin\theta + E'_y \cos\theta$$

substituting for E'_x and $E'_y \propto \cos\omega t$ and $\sin\omega t$,

$$E_x = E_0 \cos(\omega t + \theta),$$
$$E_y = E_0 \sin(\omega t + \theta)$$

These describe circularly polarised light with a phase change of θ. Changing the sign of E_y is equivalent to reflecting the electric vector in the x−axis. This changes the

sense of circular motion.

10.31 This is just a restatement of the previous problem. Since the E_y components have opposite signs for opposite circular polarisations, they cancel, leaving linear polarisation along x. If we want linear polarisation along x', we should use $E'_x \propto \cos\omega t, E'_y \propto \pm\sin\omega t$ to build the two circular waves. Coming back to x and y components, one circularly polarised wave is shifted in phase by $+\theta$ and other by $-\theta$. The rotation of linearly polarised waves by sugar solution can be thought of as a difference in refractive index between the two opposite circular waves, producing a phase difference between them.

10.32 The angle between successive polaroids is $\pi/2N$. The fractional intensity is $[\cos\{(\pi/2N)\}]^{2N}$. Trying out larger values of N will quickly convince you that this approaches 1 for large N! A mathematical expression for large N requires (i) an approximate expression for $\cos\theta$ for small θ (ii) the definition of the exponential function. If you know these, you can write, for large N,

$$\left[1 - \frac{1}{2}\left(\frac{\pi}{2N}\right)^2\right]^{2N} \approx \exp\left(-\frac{\pi^2}{4N}\right)$$

10.33 (a) Changing the sign of E_y relative to E_x reflects the polarisation in the $x-$axis, we get linear polarisation along $-\theta$.

(b) The sense of circular polarisation is reversed.

10.34 The visibility of the fringes is poorest when the path difference p is an integral multiple of λ_1 and a half integral multiple of λ_2 (for example). As p is increased, this happens first when

$$\frac{p}{\lambda_1} - \frac{p}{\lambda_2} = \frac{1}{2};$$

$$p = \frac{1}{2}\left(\frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1}\right)$$

$$= 0.29\text{mm}.$$

10.35 Divide the single slit into n smaller slits of width $a' = a/n$. The angle $\theta = n\lambda/a = \lambda/a'$. Each of the smaller slits sends zero intensity in the direction θ. The combination gives zero intensity as well.

10.36 The diffraction angle λ/a causes a spreading of $(L\lambda/a)$ in the size of the spot. This becomes large when a is small. Adding the two kinds of spreading (for simplicity, this is not strictly true!) we get a spot size $a+(L\lambda/a)$. To find the minimum value of this, write it as $\sqrt{[a-(L\lambda/a)]^2 + 4L\lambda}$. The minimum is when $a = (L\lambda/a)$ i.e. the geometric and diffraction broadening are equal. The minimum value is $2\sqrt{L\lambda}$.

10.37 Let the beams make angles $\pm(1/2)\theta$ with the normal to the screen. Moving parallel to the screen by a distance x brings one closer to the source of one beam by $x\theta/2$ and further from the source of the other by $x\theta/2$. (Draw a figure!) This causes a change in $x\theta$ in the path difference. Equating this to λ gives

$x = \lambda/\theta$ for the spacing of two bright fringes. Using the film as a grating in first order gives a change in direction of $(\lambda/x) = (\lambda/(\lambda/\theta)) = \theta$. This was the angle between the original two beams. Illuminating the film with one of the two beams used to make it "brings out" the other, inclined at θ. This problem illustrates the principle of holography. Light from a laser illuminates both, some object and a photographic film. The light scattered from the object interferes with the direct light and gives a permanent record on the film. The developed film is used as a grating using the laser beam alone. The wavefront emerging from the grating (called a hologram) includes a copy of the wave scattered by the original object. To the eye, it presents the full three dimensional appearance of that object. The person who first suggested this technique, D. Gabor, was awarded the Nobel prize for physics.

Chapter 11

11.1 $(I/(70)^2) = (60/(105)^2)$ which gives $I = 26.7$ cd.

11.2 Because of inverse-square dependence on distance, light energy falling per second reduces by a factor of 4. Therefore, to receive the same amount of light, exposure time should be increased to $2.5 \times 4 = 10$s.

11.3 Estimated speed of light $= (1.5 \times 10^{11} \times 2)/990 = 3.03 \times 10^8$ ms^{-1}

11.4 Time for the round trip travel by light $= (2 \times 10^4/c)$ s. Angle between the centre of a tooth and the centre of the gap next to it $360°/(2 \times 540) = (1/3)°$. Time required by the wheel to rotate by $(1/3)° = 0.072/(360 \times 3)$s. Equating the two times, $c = 3.0 \times 10^8$ ms^{-1}.

11.5 Time required by light for the round trip (prism to mirror and back)$=$ $(9.0 \times 10^4/c)$s. During this time the prism makes one eighth of a revolution. Time required for 1/8 rev $=$ $1/(8 \times 416.7)$ s. Equating the two times, we get

$c = 3.00 \times 10^8$ms^{-1}.

11.7 The screen should be placed 54 cm. from the mirror. The image is real, inverted and magnified. The size of the image is 5.0 cm. If the candle is moved closer, the screen would have to be moved farther and farther. Closer than 18 cm. from the mirror, the image gets virtual and can not be collected on the screen.

11.8 Virtual image located 6.7 cm behind the mirror. Magnification $= 5/9$, i.e., the size of the image is reduced to $(5/9) \times 4.5 = 2.5$ cm. As the needle is moved farther from the mirror, the image moves towards the focus (but never beyond) and gets progressively diminished in size.

11.9 Image located at 16 (2/3) cm. Magnification has a magnitude of 2/3. The image of the wire is a square of side $(2/3) \times 3 = 2.0$ cm., i.e. area 4.0 cm^2.

11.10 1.33; $(9.4 - 7.7) = 1.7$ cm

11.11

$$^a n_g = \frac{\sin 60°}{\sin 35°} = 1.51$$

$$^a n_w = \frac{\sin 60°}{\sin 41°} = 1.32$$

$$\frac{\sin 45°}{\sin r} = {}^w n_g = \frac{^a n_g}{^a n_w} = 1.144$$

which gives $\sin r = 0.6181$ i.e., $r \simeq 38°$

11.12 If r is the radius (in m) of the largest circle from which light comes out and i_c is the critical angle for water-air interface, $r = 0.8 \times \tan i_c$ and $\sin i_c = 1/1.33 \simeq 0.75$

$$\text{Area} = \frac{\pi \times (0.8)^2 \times (0.75)^2}{1 - (0.75)^2}$$

$$= 2.6\text{m}^2$$

11.13
$$n = \frac{\sin[(A + d_m)/2]}{\sin[A/2]}$$

$$= \frac{\sin 50°}{\sin 30°}$$

$$= 2 \times 0.766 = 1.532 \simeq 1.53$$

$$\frac{\sin(30° + d'_m/2)}{\sin 30°} = \frac{1.53}{1.33}$$

which gives $d'_m \simeq 10°$

11.16 Image formed on a screen is real. The lens must be a converging lens. From the lens equation, $f = 30$ cm. Size of the image = 10 cm.

11.17 Use the lensmaker's formula to obtain $R = 22$ cm.

11.18 Here the object is virtual and the image is real. $u = +12$ cm. (object on right; virtual).

(a) $f = +20$ cm

$$\frac{1}{v} = \frac{1}{f} + \frac{1}{u} = \frac{1}{20} + \frac{1}{12}$$

$$= \frac{2}{15} \quad \text{i.e.} \quad v = 7.5 \text{cm}.$$

(image on right; real). 7.5 cm. from the lens.

(b) $f = -16$ cm

$$\frac{1}{v} = -\frac{1}{16} + \frac{1}{12} = \frac{1}{48}$$

i.e. $v = 48$ cm. (image on right; real). 4.8 cm. from the lens.

11.19 Image is erect, virtual and located 8.4 cm. from the lens on the same side as the object. It is diminished to a size = (8.4/14) × 3 = 1.8 cm. As the object is moved away from the lens, the virtual image moves towards the focus of the lens (but never beyond), and progressively diminishes in size. Note, when the object is placed at the focus of the concave lens (21 cm.), the image is located at 10.5 cm. (not at infinity as one might wrongly think). A *virtual* object at the focus of a concave lens produces an image at infinity.

11.20 (a) The image due to the first lens becomes an object for the second lens. Use this fact together with the lens equation for each lens to obtain the required relation.

(b) A diverging lens of focal length 60 cm.

11.21 (b) 6

11.22 (a) $v_e = -25$ cm, $f_e = 6.25$ cm give $u_e = -5$ cm

$v_0 = 15 - 5 = 10$ cm,

$f_0 =$ $u_0 = -2.5$ cm

Magnifying power

$$= \frac{10}{2.5} \times \frac{25}{5} = 20$$

(b) $u_e = -6.25$ cm, $v_0 = 15 - 6.25 = 8.75$ cm, $f_0 = 2.0$ cm. Therefore, $u_0 = -(70/27) = -2.59$ cm.

Magnifying power

$$= \frac{v_0}{|u_0|} \times (25/6.25)$$

$$= \frac{27}{8} \times 4 = 13.5$$

11.23 Angular magnification of the eyepiece for image at 25 cm

$$= \frac{25}{2.5} + 1 = 11;$$

$$|u_e| = \frac{25}{11}\text{ cm} = 2.27\text{ cm}.$$

Now $\frac{1}{v_0} + \frac{1}{0.9} = (1/0.8)$

i.e. $v_0 = 7.2$ cm

separation $= 7.2 + 2.27 = 9.47$ cm

Magnifying power

$$= 11 \times \frac{7.2}{0.9} = 88$$

11.24 (b) 24; 150 cm

11.25 (a) Angular magnification

$$= \frac{15}{0.01} = 1500$$

(b) If d is the diameter of the image (in cm),

$$\frac{d}{1500} = \frac{3.48 \times 10^6}{3.8 \times 10^8}$$

i.e. $d = 13.7$ cm

11.26 Data on focal lengths not needed. The resolving power of the telescope is determined by

$$\theta = \frac{1.22\lambda}{D} = \frac{1.22 \times 6 \times 10^{-7}}{0.60}$$

$$= 1.22 \times 10^{-6}\text{ rad}$$

where the value of λ chosen corresponds roughly to the wavelength of yellow light. Now the transverse separation between the two sources subtends an angle equal to

$$\frac{10^{10}\text{m}}{9.46 \times 10^{19}\text{m}} \simeq 10^{-10}\text{rad},$$

where we have used 1 light year = 9.46×10^{15}m. This angle is much too small compared to θ above. The two stars of the binary can not be resolved by the given telescope.

11.27 (a) 5

(b) Distant images arise due to multiple reflections. At each reflection, part of the incident intensity of light is lost due to absorption etc.

(c) For searchlight, a parallel beam of light is required. A divergent beam is not useful because then intensity diminishes with distance. (recall inverse - square dependence.) Now in case of a concave spherical mirror only paraxial rays (i.e., those close to the axis) parallel to the axis are brought to a sharp focus. Conversely, a source placed at the focus of a concave mirror produces a parallel beam only close to the axis. Rays reflected from points not close to the axis give rise to a divergent beam. This is not so in case of a parabolic mirror. A source placed at the focus of a parabolic mirror produces a parallel beam of wide cross-section. Hence its use as a searchlight mirror.

(d) A convex mirror gives a much wider field of view of the traffic at your back, than a plane mirror of the same size. However it gives an erroneous idea of the movement of the vehicles. Because of the first advantage it is still preferred to a plane mirror.

11.28 (a) Focal length of a mirror is

about half its radius of curvature and has nothing to do with the external medium. The focal length of the convex lens will increase because the refractive index of glass with respect to water is less than refractive index of glass with respect to air.

(b) The air layers closer to the ground are hotter than higher layers. Oblique rays coming from distant sky therefore travel from denser to rarer parts of the atmosphere and get more and more oblique. When the angle of incidence exceeds critical angle (for dense air-rarer air interface), rays get totally reflected and may enter the observer's eye. The observer therefore sees a reflected image of the distant parts of the sky.

(c) The apparent position of a star is slightly different from the actual position due to refraction of starlight by the atmosphere. Further, this apparent position is not stationary, since the conditions of the refracting medium are not stationary. Starlight travels through fluctuating masses of air in motion with changing conditions of temperature, temperature gradients etc. The fluctuating apparent position of the star give rise to the twinkling effect.

(d) Since the atmosphere bends starlight towards the normal, the apparent position of a star is slightly 'above' its actual position. Thus even when the sun has actually set (i.e. gone below the horizon) its apparent position remains above the horizon for some time.

11.29 (a) A white body reflects all the light incident on it. A black body, in contrast absorbs all of the light energy incident on it which then gets converted into heat. This is why dark dresses are used to keep warm in winter.

(b) If an object looks blue in white light, it means it absorbs all the colours except those in the blue region. Light from a sodium lamp is yellow which therefore is nearly wholly absorbed by the object. The object will appear black.

(c) The mask has a filter that absorbs the ultraviolet radiation (which is dangerous for eyes) produced by the welding arc.

(d) Scattering of light by the atmosphere is colour-dependent. Blue light is scattered much more strongly than red light. The blue component of sunlight is therefore proportionately more in light coming from different parts of the atmosphere. This gives the impression of a blue sky. In the evening when the sun is near the horizon, sunlight has to travel through much greater distance than at noon. Thus a much larger proportion of the blue component of sunlight

gets scattered away. Light reaching the observer therefore has a larger proportion of the remaining colours. The sun therefore appears orange or red.

11.30 (a) The camera lens is moved towards or away from the film. For instance, if a very distant object is being focused, the distance between the lens and the film is about the focal length of the lens. At closer object distance, the camera lens must be moved away from the film to provide the required greater image distance. In practice, because of the small focal length of the lens in an ordinary camera (about 5 cm) and the large object distance involved, only a minor movement of the camera lens serves the purpose.

(b) f-number of a camera lens describes the ratio of focal length to the diameter of the aperture of the lens. Thus $f/11$ means the diameter of the aperture is focal length divided by 11. Now the capacity to collect light depends on the square of the aperture size. Thus the given sequence of apertures (for a fixed focal length) have a light gathering capacity proportional to

$$\frac{1}{2^2}, \frac{1}{(2.8)^2}, \frac{1}{4^2}, \frac{1}{(5.6)^2}, \frac{1}{8^2}, \frac{1}{11^2}$$

which is roughly in the ratio

$$1 : \frac{1}{2} : \frac{1}{4} : \frac{1}{8} : \frac{1}{16} : \frac{1}{32}$$

The corresponding exposure times required to receive the same total amount of light are in the ratio: 2:4:8:16:32.

(c) The shutter controls the exposure time, usually in convenient steps such as (1/60)s, (1/120)s etc. For a proper photograph, the total amount of light recieved by the film should be within certain limits. The aperture size and the exposure time together determine the total amount of light received.

Answers to Additional Exercises

11.31 The reflected rays gets deflected by twice the angle of rotation of the mirror. Therefore, $d/1.5 = \tan 7°$ i.e., $d = 18.4$ cm.

11.32 For the head to be seen by the eye, the top edge of the mirror should not be lower than 1.44 m from the ground. For the foot to be seen, the bottom edge of the mirror should not be higher than 0.69 m from the ground. Thus the minimum length of the mirror for a full view is $(1.44-0.69) = 0.75$ m. This minimum length is the same for any level, but the positions of the top and bottom edges of the mirror will depend on the eye-level.

11.33 Time required by light for the round trip M to N to M

$$= \frac{4.5 \times 10^{-4}}{3 \times 10^{8}} = 1.5 \times 10^{-4}\text{s}$$

During this time the mirror rotates by $(1/2) \times 27 = 13.5°$. Speed of ro-

tation of the mirror

$$= \frac{13.5}{360} \times \frac{1}{10^{-4} \times 1.5} \text{rev/s}$$
$$= 250 \text{rev s}^{-1}.$$

11.34 (a) $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ or $\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$

$f < 0$ (concave mirror);

$u < 0$ (object on left)

For $2f < u < f$ implies

$$\frac{1}{2f} > \frac{1}{u} > \frac{1}{f}$$

$$\text{or} \quad -\frac{1}{2f} < -\frac{1}{u} < -\frac{1}{f}$$

$$\text{or} \quad \frac{1}{f} - \frac{1}{2f} < \frac{1}{f} - \frac{1}{u} < 0$$

$$\text{or} \quad \frac{1}{2f} < \frac{1}{v} < 0$$

which means $v < 0$ (image on left, real) the image lies beyond $2f$. The image is real because v is negative.

(b) $\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$.

Now for a convex mirror $f > 0$. Also we have $u < 0$ (object on left).

Therefore, $(1/v)$ or $v > 0$ (image on right; virtual), i.e., the image is virtual whatever be the value of u.

(c) $\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$.

Since $f > 0$ (convex mirror) and $u < 0, (1/v) > (1/f)$ i.e., $v < f$ (image located between the pole and the focus). And from above $v < |u|$ (image diminished).

(d) $\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$

$f < 0$ (concave mirror), $f < u < 0$, implies

$$\frac{1}{f} - \frac{1}{u} > 0,$$

i.e., $(1/v) > 0$ or $v > 0$ (image on right; virtual).

$$\text{Also} \quad \frac{1}{v} < \frac{1}{|u|}$$

i.e., $v > |u|$ (image enlarged).

11.35 The pin appears raised by 15 $(1 - (1/15)) = 5.0$ cm. You can see from an explicit ray diagram that the answer is independent of the location of the slab (for small angles of incidence).

11.36 Use the important result of 11.35. The displacement between the virtual image due to refraction and the object depends only on the thickness and refractive index of the intervening medium. If a second medium is interposed, the image due to the first becomes an object for the second medium. The total displacement between the final image and the object is then the sum of displacements due to each medium. Thus total displacement

$$= 4\left(1 - \frac{1}{1.5}\right) + 6\left(1 - \frac{1}{1.4}\right) + 8\left(1 - \frac{1}{1.3}\right) = 4.9\text{cm}.$$

11.37 You will see that rays inside the prism undergo total internal reflections, twice (once on each side of the prism) in case (a); and once (on the hypotenuse of the prism) in case (b), provided the refracted ray in the prism meets the hypotenuse.

11.38 (a) $\sin i_c = 1.44/1.68$ which gives $i'_c = 59°$. Total internal reflection takes place when $i' > 50°$ or when $r < r_{\max} = 31°$.

Now $(\sin i_{\max}/\sin r_{\max}) = 1.68$ which gives $i_{max} \simeq 60°$. Thus all incident rays of angles in the range $0 < i < 60°$ will suffer total internal reflections in the pipe. (If the length of the pipe is finite, which it is in practice, there will be a lower limit on i determined by the ratio of the diameter to the length of the pipe).

(b) If there is no outer coating, $i'_c = \sin^{-1}(1/1.68) = 36.5°$. Now, $i = 90°$ will have $r = 36.5°$ and $i' = 53.5°$ which is greater than i'_c. Thus all incident rays (in the range $0 < i < 90°$) will suffer internal reflections.

11.39 The slit can be viewed by the light reflected from either face of the prism. For two parallel incident rays, one on each face, the angle between the corresponding reflected rays can be shown to be twice the angle of the prism. Therefore $A = 72°$.

11.40 Lens equation

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

(New cartesian convention).

(a) Convex lens $f > 0$; $u < 0$ (object on left).

For $0 < |u| < f$

$$\frac{1}{v} = \frac{1}{f} + \frac{1}{u}$$
$$= \frac{1}{f} - \frac{1}{|u|} < 0$$

i.e., $v < 0$

(image on left; virtual). Also for this case

$$\frac{1}{|v|} < \frac{1}{|u|} \quad \text{i.e.,} \quad |v| > |u|$$

(image enlarged)

(b) Concave lens: $f < 0; u < 0$ (object on left)

$$\frac{1}{v} = \frac{1}{f} + \frac{1}{u}$$
$$= -\left(\frac{1}{|f|} + \frac{1}{|u|}\right) < 0$$

for all u i.e., $v < 0$ for all u (image on left; virtual). Also

$$\frac{1}{|v|} = \frac{1}{|f|} + \frac{1}{|u|}$$

i.e., $\dfrac{1}{|v|} > \dfrac{1}{|u|}$

i.e., $|v| < |u|$

(image diminished).

11.41 (a) The front surface of a thick glass mirror is both reflecting and refracting. The back surface is silvered and acts as a mirror. Images arise due to reflection of the incident light by the front surface then by the back surface (the bright image), followed by multiple reflections of light within the glass by the front and back surfaces. You can draw a ray diagram to see all this more clearly.

(b) You read because of the light scattered by the newspaper, i.e., the newspaper causes diffuse reflection. An image is seen only when reflection is regular (or specular as it is sometimes called), i.e., when parallel incident rays are reflected in parallel directions. The surface inhomogeneities of the paper are responsible for diffuse reflection

(c) Rays converging to a point 'behind' a plane or convex mirror are reflected to a point in front of the mirror on a screen. In other words a plane or convex mirror can produce a real image if the object is virtual. Convince yourself by drawing an appropriate ray diagram.

(d) Nature exhibits left-right symmetry. That is, physical laws are identical for you and your mirror image. That is, the movements of your mirror image are perfectly physically possible movements and if an outsider views the two films he cannot say which of the two is the mirror image film. Of course, any additional left-right *asymmetric* initial information which you may have can help decide which is which. For example, if it is known that you are say a left-hander, then your image will appear to be a right-hander. And from watching the two films you can tell which is the mirror image film. The important thing is not to confuse this possibility of asymmetric initial condition with left-right asymmetry in physical laws, which as already said, do not distinguish between you and your mirror image. (Note: The left-right symmetry of nature, however, does not hold good at the sub-atomic level for some types of processes).

11.42 (a) When the reflected or refracted rays are divergent, the image is virtual. The divergent rays can be converged on to a screen by means of an appropriate converging lens. The convex lens of the eye does just that. The virtual image here serves as an object for the lens to produce a real image. Note, the screen here is not located at the position of the virtual image. There is no contradiction.

(b) Taller

(c) The apparent depth for oblique viewing decreases from its value for near-normal viewing. Convince yourself of this fact by drawing ray diagrams for different positions of the observer.

(d) Refractive index of diamond is about 2.42, much larger than that of ordinary glass (about 1.5). The critical angle for diamond is about 24°, much less than that of glass. A skilled diamond cutter exploits the larger range of angles of incidence (in the diamond), 24° to 90°, to ensure that light entering the diamond is totally

reflected from many faces before getting out – thus producing a sparkling effect.

11.43 For a fixed distance s between object and screen, the lens equation does not give a real solution for u or v if f is greater than $s/4$. Therefore $f_{\max} = 0.75$ m.

11.44 (a) $-u$ and v interchange their values at the two locations

$$-u \text{ (or } v) = \frac{90-20}{2} = 35\,\text{cm}$$

$$v \text{ (or } -u) = 55\,\text{cm}.$$

$$f = \frac{55 \times 35}{90} = 21.4\,\text{cm}.$$

(b) $\dfrac{4.6}{h} = \dfrac{v}{|u|}$,

$\dfrac{1.7}{h} = \dfrac{|u|}{v}$

(because $|u|$, v interchange their values at the two locations).

Thus $h = \sqrt{4.6 \times 1.7} = 2.8$ cm.

11.45 (a) (i) Let a parallel beam be incident from the left on the convex lens first.

$f_1 = 30$ cm, $u_1 = -\infty$, give $v_1 = +30$ cm. This image becomes a virtual object for the second lens.
$f_2 = -20$ cm, $u_2 = +(30 - 8) = +22$ cm. Therefore,

$$\frac{1}{v_2} = -\frac{1}{20} + \frac{1}{22}$$

which gives $v_2 = -220$ cm.
The parallel incident beam appears to diverge from a point 216 cm from the centre of the two-lens system.

(ii) Let the parallel beam be incident from the left on the concave lens first:
$f_1 = -20$ cm, $u_1 = -\infty$, give $v_1 = -20$ cm. This image becomes a real object for the second lens:
$f_2 = +30$ cm, $u_2 = -(20 + 8) = -28$ cm. Therefore,

$$\frac{1}{v_2} = \frac{1}{30} - \frac{1}{28}$$

i.e., $v_2 = -420$ cm.

The parallel incident beam appears to diverge from a point 416 cm on the left of the centre of the two-lens system.
Clearly, the answer depends on which side of the lens system the parallel beam is incident. *Further, we do not have a simple lens equation true for all u (and v) in terms of a definite constant of the system (the constant being determined by f_1 and f_2, and the separation between the lenses).* The notion of effective focal length therefore, does not seem to be meaningful for this system

(b) $u_1 = -40$ cm., $f_1 = 30$ cm. give $(1/v_1) + (1/40) = (1/30)$
i.e., $v_1 = 120$ cm.

Magnitude of magnification due to the first (convex) lens

$= 120/40 = 3$.

$$u_2 = +(120-8) = +112\text{cm}$$
(object virtual);

$$f_2 = -20 \text{ cm which give}$$

$$\frac{1}{v_2} = -\frac{1}{20} + \frac{1}{112}$$

$$\text{i.e., } v_2 = -\frac{112 \times 20}{92} \text{ cm}$$

Magnitude of magnification due to the second (concave) lens = 20/92.

Net magnitude of magnification = 3 × (20/92) = 0.652.

Size of the image = 0.652 × 1.5 = 0.98 cm.

11.46 If the refracted ray in the prism is incident on the second face at the critical angle i_c, the angle of refraction r at the first face is $60° - i_c$.

Now $i_c = \sin^{-1}(1/1.524) \simeq 41°$.

Therefore $r = 19°$.

$$\sin i = (\sin 90°) \times 1.524 = 0.4962$$
$$i = \sin^{-1} 0.4962 \simeq 30°.$$

11.47 The first measurement gives the focal length f of the combination of the convex lens and the plano-convex liquid lens. The second measurement gives the focal length f_1 of the convex lens. The focal length f_2 of the plano-convex lens is then given by:

$$\frac{1}{f_2} = \frac{1}{45} - \frac{1}{30} = -\frac{1}{90}$$

$$\text{i.e., } f_2 = -90\text{cm}.$$

Using the lens maker's formula for the equiconvex lens,

$$\frac{1}{30} = (1.5-1)\left(\frac{1}{R} + \frac{1}{R}\right)$$

which gives $R = 30$ cm.

The same formula applied to the plano-convex lens gives

$$-\frac{1}{90} = (n-1)\left(\frac{1}{30} + 0\right)$$

from which the refractive index of liquid, $n = 1.33$.

11.48 Use

$$n = \frac{\sin[(A+\delta_{\text{min}})/2]}{\sin[A/2]}$$

to get

Crown glass:

$$n_b = 1.520, n_r = 1.509, n_y = 1.516$$

Flint glass:

$$n_b = 1.666, n_r = 1.644, n_y = 1.657$$

Use

$$\omega = \frac{n_b - n_r}{n_y - 1}$$

$\omega = 0.0213$ crown glass

$\omega = 0.0335$ flint glass.

The ratio of dispersive power of flint glass to crown glass is 1.57.

11.49 Two identical prisms made of the same material placed with their bases on opposite sides (of the incident white light) and faces touching (or parallel) will neither deviate nor disperse, but will merely produce a parallel displacement of the beam.

(a) To deviate without dispersion, choose say the first prism to be of crown glass, and take for the

second prism a flint prism of suitably chosen refracting angle (smaller than that of crown glass prism because the flint prism disperses more) so that dispersion due to the first is nullified by the second.

(b) To disperse without deviation, increase the angle of the flint glass prism (i.e., try flint glass prisms of greater and greater angle) so that deviations due to the two prisms are equal and opposite. (The flint glass prism angle will still be smaller than that of crown glass because flint glass has higher refractive index than that of crown glass). Because of the adjustments involved for so many colours, these are not meant to be precise arrangements for the purposes required.

11.50 We can take the average focal length of 15 cm to correspond to yellow colour. At a point slightly less than 15 cm, the centre of the spot will be violet with a red edge and other colours in between. As the screen is moved away, the centre of the spot changes through a succession of blue, green, yellow, orange and finally red. When the centre of the spot is red, the edge will be violet with other colours in between.

11.51 (a) Using the lensmaker's formula for blue, red and yellow separately, we obtain:

$$\frac{1}{f_1^b} - \frac{1}{f_1^r} = \frac{\omega_1}{f_1^y}$$

and

$$\frac{1}{f_2^b} - \frac{1}{f_2^r} = \frac{\omega_2}{f_2^y}$$

Thus

$$\begin{aligned}\frac{1}{f^b} - \frac{1}{f^r} &= \left(\frac{1}{f_1^b} + \frac{1}{f_2^b}\right) \\ &\quad - \left(\frac{1}{f_1^r} + \frac{1}{f_2^r}\right) \\ &= \frac{\omega_1}{f_1^y} + \frac{\omega_2}{f_2^y}\end{aligned}$$

For an 'achromatic doublet', $f^b = f^r$ which gives

$$\frac{f_1^y}{f_2^y} = -\frac{\omega_1}{\omega_2}$$

(b) Combine the given flint glass lens ($f_1^y = 15$ cm) with a concave lens made of crown glass of focal length $f_2^y = -10$ cm.

11.52 Double convex lens (crown glass)

$$\begin{aligned}\frac{1}{f_b} - \frac{1}{f_r} &= \frac{(n_b - n_r) \times 2}{15} \\ &= (1.520 - 1.509) \times \frac{2}{15} \\ &= 0.0014666\text{cm}^{-1}\end{aligned}$$

Flint glass lens

$$\begin{aligned}\frac{1}{f_b} - \frac{1}{f_r} &= (1.660 - 1.644) \\ &\quad \times \left(-\frac{1}{15} + \frac{1}{R}\right) \\ &= -0.0014666\text{cm}^{-1}\end{aligned}$$

This gives $R = \infty$

The other surface of the flint glass lens is plane. It is a plano-convex lens.

11.53 (a) Not necessarily. A material may reflect one colour strongly

and transmit another colour e.g., some lubricating oils reflects green light and transmit red light.

(b) Green with a tinge of blue and yellow.

11.54 To see objects at infinity, the eye uses its least converging power = 40 + 20 = 60 dioptres. This gives a rough idea of the distance between the retina and the cornea – eyelens: (5/3) cm. To focus an object at the near point ($u = -25$ cm) on the retina ($v = 5/3$ cm), the focal length should be

$$\left[\frac{1}{25}+\frac{3}{5}\right]^{-1} = \frac{25}{16}\text{ cm}$$

corresponding to a converging power of 64 dioptres. The power of the eyelens then is 64 – 40 = 24 dioptres. The range of accomodation of the eylens is roughly 20 to 24 dioptres.

11.55 No, a person may have normal ability of accomodation of the eyelens and yet may be myopic or hyperopic. Myopia arises when the eyeball from front to back gets too elongated; hyperopia arises when it gets too shortened. In practice, in addition the eyelens may also lose some of its ability of accomodation. When the eyeball has normal length but the eyelens loses partially its ability of accomodation (as happens with increasing age for any normal eye) the 'defect' is called presbyopia and is corrected in the same manner as hyperopia.

11.56 (a) Concave lens of focal length = –80 cm i.e. of power = –1.25 dioptres.

(b) No. The concave lens in fact reduces the size of the object (image distance less than object distance), but the angle subtended by the distant object at the eye is the same as the angle subtended by the image (on the far point) at the eye. The eye is able to see distant objects not because the corrective lens magnifies the object but because it brings the object (i.e. it produces virtual image of the object) at the far point of the eye which then can be focussed by the eyelens on the retina.

(c) The myopic person may have a normal near point i.e. about 25 cm. (or even less). In order to read a book with his spectacles (for distant vision), he must keep the book at a greater distance than 25 cm so that the image of the book by the concave lens is produced not closer than 25 cm. The angular size of the book (or its image) at the greater distance is evidently less than the angular size when the book is placed at 25 cm and no spectacles are used. Hence the person prefers to remove his spectacles while reading.

11.57 (a) $u = -25$ cm, $v = -75$ cm

$$\frac{1}{f} = \frac{1}{25} - \frac{1}{75}$$

i.e., $f = 37.5$cm.

The corrective lens has a converging power of + 2.67 dioptres.

(b) The corrective lens produces a virtual image (at 75 cm) of an object at 25 cm. The angular size of this image is the same as that of the object. In this sense the lens does not magnify the object but merely brings the object to the near point of the eye which then gets focussed by the eyelens on the retina. However the angular size is greater than that of the same object at the near point (75 cm) viewed without the spectacles.

(c) A hyperopic eye may have normal far point i.e., it may have enough converging power to focus parallel rays from infinity on the retina of the shortened eyeball. Wearing spectacles of converging lenses (used for near vision) will amount to more converging power than needed for parallel rays. The result will be that distant objects may get focussed in front of the retina (like in a myopic eye) and will appear blurred.

11.58 (a) $f = -(1/0.8)$ m $= -125$ cm; $v = -80$ cm.

Using the lens equation, $u = -222$ cm. He can see objects up to 2.22 m clearly.

(b) $f = +1/1$ m $= 100$ cm; $v = -75$ cm. Using the lens equation, $u = -42.9$ cm. His nearest distance of distinct vision is about 43 cm.

11.59 The far point of the person is 100 cm, while his near point may have been normal (about 25 cm). Objects at infinity produce virtual images at 100 cm (using spectacles). To view closer objects i.e., those which are (or whose images using the spectacles are) between 100 cm and 25 cm, the person uses the ability of accomodation of his eyelens. This ability usually gets partially lost in old age(presbyopia). The near point of the person receds to 50 cm. To view objects at 25 cm clearly, the person needs converging lens of power +2 dioptres.

11.60 This defect (called astigmatism) arises because the curvature of the cornea plus eyelens-refracting system is not the same in different planes. [The eyelens is usually spherical i.e., has the same curvature in different planes but the cornea is not spherical in case of an astigmatic eye]. In the present case, the curvature in the vertical plane is enough, so sharp images of vertical wires can be formed on the retina. But the curvature is insufficient in the horizontal plane, so horizontal wires appear blurred. The defect can be corrected by using a cylindrical lens with its axis along the vertical. Clearly, parallel rays in the vertical plane will suffer no extra refraction, but those in the horizontal plane can get the required extra convergence due to refraction by the curved surface of the cylindrical lens if the curvature of the cylindrical surface is chosen appropriately.

11.61 (a) Closest distance =

$4\frac{1}{6}$ cm $\simeq 4.2$ cm

$\left(\text{use:} \quad -\frac{1}{25} - \frac{1}{u} = \frac{1}{5}\right)$

Farthest distance = 5 cm

$\left(\text{use:} \quad -\frac{1}{\infty} - \frac{1}{u} = \frac{1}{5}\right)$

(b) Maximum angular magnification = $[25/(25/6)] = 6$

Minimum angular magnification = $(25/5) = 5$

11.62 (a) $\frac{1}{v} + \frac{1}{9} = \frac{1}{10}$

i.e. $v = -90$ cm.

Magnitude of magnification = $90/9 = 10$.

Each square in the virtual image has an area $10 \times 10 \times 1 = 100 \text{ mm}^2 = 1 \text{ cm}^2$.

(b) Magnifying power = $25/9 = 2.8$

(c) No, magnification of an image by a lens and angular magnification (or magnifying power) of an optical instrument are two separate things. The latter is the ratio of the angular size of the object (which is equal to the angular size of the image even if the image is magnified) to the angular size of the object if placed at the near point (25 cm). Thus magnification magnitude is $|(v/u)|$ and magnifying power is $(25/|u|)$. Only when the image is located at the near point $|v| = 25$ cm are the two quantities equal.

11.63 (a) Maximum magnifying power is obtained when the image is at the near point (25 cm)

$$-\frac{1}{25} - \frac{1}{u} = \frac{1}{10}$$

$$\text{i.e.,} \quad u = -\frac{50}{7} = -7.14\text{cm}.$$

(b) Magnitude of magnification = $(25/|u|) = 3.5$.

(c) Magnifying power = 3.5

Yes, the magnifying power (when the image is produced at 25 cm) is equal to the magnitude of magnification. See answer to 11.62.

11.64 Magnification = $\sqrt{(6.25/1)} = 2.5$

$$v = +2.5u$$

$$+\frac{1}{2.5u} - \frac{1}{u} = \frac{1}{10}$$

$$\text{i.e.,} \quad u = -6\text{cm},$$

$$|v| = 15\text{cm}.$$

The virtual image is closer than the normal near point (25 cm) and cannot be seen by the eye distinctly.

11.65 (a) Even though the absolute image size is bigger than the object size, the angular size of the image is equal to the angular size of the object. The magnifier helps in the following way: without it object would be placed no closer than 25 cm; with it the object can be placed much closer. The closer object has larger angular size than the same object at 25 cm. It is in this sense that angular magnification is achieved.

(b) Yes, it decreases a little because the angle subtended at the eye is then slightly less than the angle subtended at the lens. The effect is negligible if the image is a very large distance away. [Note: when the eye is separated from the lens, the angles subtended at the eye by the first object and its image are not equal.]

(c) First, grinding lens of very small focal lengths is not easy. More important, if you decrease focal length, aberrations (both spherical and chromatic) become more pronounced. So in practice you can't get a magnifying power of more than 3 or so with a simple convex lens. However using an aberration corrected lens system, one can increase this limit by a factor of 10 or so.

(d) Angular magnification of eye-piece is $(25/f_e)+1$ (f_e in cm) which increases if f_e is smaller. Further, magnification of the objective is given by

$$\frac{v_0}{|u_0|}=\frac{1}{(|u_0|/f_0)-1}$$

which is large when $|u_0|$ is slightly greater than f_0. Now the microscope is used for viewing very close objects. So $|u_0|$ is small, and so is f_0.

(e) The image of the objective in the eye-piece is known as 'eye-ring.' All the rays from the object refracted by objective go though the eye-ring. Therefore it is an ideal position for our eyes for viewing. If we place our eyes too close to the eye-piece, we shall not collect much of the light and also reduce our field of view. If we position our eyes on the eye-ring and the area of the pupil of our eye is greater or equal to the area of the eye ring our eyes will collect all the light refracted by the objective. The precise location of the eye ring naturally depends on the separation between the objective and the eye-piece. When you view through a microscope by placing your eyes on one end, the ideal distance between the eye and the eye-piece is usually built in the design of the instrument.

11.66 Assume microscope in normal use i.e., image at 25 cm.

Angular magnification of the eye-piece

$$=\frac{25}{5}+1=6.$$

Magnification of the objective

$$=\frac{30}{6}=5$$

$$\frac{1}{5u_0}-\frac{1}{u_0}=\frac{1}{1.25}$$

which gives $u_0=-1.5$ cm; $v_0=$ 7.5 cm . $|u_e|=(25/6)=4.17$ cm.

The separation between the objective and the eye-piece should be 7.5 + 4.17 = 11.67 cm. Further the object should be placed 1.5 cm from the objective to obtain the desired magnification.

11.67 (a) M.P. $= (f_0/f_e) = (140/5) = 28$

(b) M.P. $= \frac{f_0}{f_e}\left[1 + \frac{f_0}{25}\right]$

$= 28 \times 1.2 = 33.6$

11.68 (a) $f_0 + f_e = 145$ cm

(b) Angle subtended by the tower $= (100/3000) = (1/30)$ rad.

Angle subtended by the image produced by the objective

$$= \frac{h}{f_0} = \frac{h}{140}.$$

Equating the two,

$$h = \frac{14}{3}\text{ cm} = 4.7\,\text{cm}.$$

(c) Magnification (magnitude) of the eye-piece

$$= \frac{25}{5} + 1 = 6$$

Height of the final image (magnitude)

$$= \frac{14}{3} \times 6 = 28\,\text{cm}.$$

11.69 $-\frac{1}{u_e} + \frac{1}{40} = \frac{1}{10}$

i.e., $|u_e| = (40/3)$ cm;

magnification of the eye-piece = 3.

Therefore diameter of the image formed by the objective $= 6/3 = 2.0$ cm

If D is the diameter of the sun (in m),

$$\frac{D}{1.5 \times 10^{11}} = \frac{2}{100}.$$

His rough estimate of the size of the sun: $D = 1.5 \times 10^9$ m (correct value $= 1.39 \times 10^9$ m)

11.70 (a) No chromatic aberration due to the objective because only reflection is involved; spherical aberration reduced by using a mirror of the shape of a paraboloid; brighter image than in a refracting telescope of equivalent size because in the latter intensity of light is partially lost due to reflection and absorption by the objective lens glass; mirror entails grinding and polishing of only one side; high resolution (as well as brightness of a point object) achieved by using a mirror of large aperture which is easier to support (its back side being available) than a lens of the same aperture.

(b) Focal length of the mirror

$= -40$ cm;

M.P. $= 40/1.6 = 25$

11.71 (a) See the answer to 11.65(e)

(b) First, let us determine the location of the eye ring: Distance of the objective from the eye-piece, $u = -(f_0 + f_e)$

Image (i.e. eye-ring) distance $= v$.

Use $\frac{1}{v} + \frac{1}{f_0 + f_e} = \frac{1}{f_e}$

to get $v = \frac{(f_0 + f_e)f_e}{f_0}$

Linear magnification

$$= \frac{v}{|u|} = \frac{f_e}{f_0}$$

But linear magnification

$$= \frac{\text{diameter of the eye ring}}{\text{diameter of the objective}}$$

Using angular magnification = (f_0/f_e), the desired result is proved.

(c) Use the result in (b). Diameter of the objective = $300 \times 0.3 =$ 90 cm

11.72 In a terrestrial telescope, the inverted image formed by the objective is made erect by positioning it at the $2f$ point of an erecting lens of focal length f. Thus the separation between the objective and the eye-piece is $f_0 + f_e + 4f = 180 + 5 + 14 = 199$ cm. Magnifying power remains unaltered: $(f_0/f_e) = (180/5) = 36$. The telescope can, of course, be used to view astronomical objects though obviously there is no need to make the 'inverted' image of a star 'upright' (This is, however, necessary for viewing a terrestrial object). The final image obtained is somewhat less bright than in an equivalent astronomical telescope because of the extra loss of some light due to reflection and absorption by the erecting lens.

11.73 (a) Draw a ray diagram for this telescope, remembering that the image formed by the objective is a virtual object for the eye-piece interposed between the objective and its focus (which is also the focus of the eye-piece in normal adjustment). The derivation is identical to that for an ordinary astronomical telescope.

(b) Separation = $150 - 7.5 = 142.5$ cm.

(c) Limited field of view because the eye cannot be positioned on the location of the eye ring between the two lenses.

11.74 Main advantage: compactness. The effective length of the telescope is three times the distance between the objective and the eye-piece.

11.75 (a) Light-gathering capacity of A is 9 time that of B.

(b) Take a star at a distance ℓ that is barely visible in a telescope using an objective of diameter d. The intensity of light from a star (of the same absolute brightness) at a distance 2ℓ is reduced by a factor of 4. To receive the same total amount of light, the diameter of the objective should be $2d$, using the result similar to (a). Thus if the distance of the star is doubled, the diameter of the objective should be doubled for the star to be again barely visible. Thus the distance a telescope can penetrate through the sky (range) is proportional to the diameter of its objective.

(c) Background is not like a point source but is like an object of finite size. If M is the magnifying power, the area of the image seen in the telescope is M^2 the area seen by the unaided eye. But M is also the ratio of the diameter of the objective to the diameter of the eye-ring.

Assuming our eye pupil fills up the eye ring, it is clear that the light received by the objective is M^2 the light received by our unaided eye. Thus, for an object of finite extension, the telescope increases the area of the image seen and the total light received in the same proportion. The net effect is that the amount of light per unit area i.e., brightness is not increased. It is, in fact, slightly reduced due to absorption etc., of some light by the various optical elements of the telescope. The argument does not go through for a point source, so that the brightness of a point source does increase using a telescope. A star, therefore appears brighter against its background when viewed through a telescope.

11.76 Linear Magnification

$$= \frac{1000}{24} = \frac{125}{3}$$

$$-\left(u + \frac{125}{3}u\right) = 12$$

$$\text{i.e.,} \quad u = \frac{-9}{32}\text{ m},$$

$$v = \frac{375}{32}\text{ m}$$

Using $\dfrac{1}{v} - \dfrac{1}{u} = \dfrac{1}{f}$,

$f = 27.5$ cm.

The projection lens should be 28.1 cm from the slide and have focal length = 27.5 cm.

11.77 (a) A strong illumination of the slide is necessary in a projector, otherwise the highly enlarged image on the screen will be very dim. If the slide is illuminated directly by a lamp, light diverging out of the outer portions of the slide will not be collected by the projecting lens (unless it is impractically big). We will then see only a small central part of the slide picture on the screen. A condensing lens helps converge light on the slide, thus enabling the whole of the slide to be imaged on the screen. Another important function of the condensing lens is explained in (c).

(b) The slide is placed very close to the condensing lens in order to utilize nearly all of the light coming out of this lens for illuminating the slide. The diagonal of the slide in 11.76 is about 4.33 cm. The diameter of the condensing lens should be obviously at least about 4.4 cm in order to illuminate the whole of the slide.

(c) An important thing in a projector is to avoid imaging the source itself on the screen. The focal length of the condensing lens should be so chosen that the image of the source is formed on the projection lens itself. Further, other parameters such as the size of the source, the diameter of the projection lens and its placement should be such that the area of the image of the source should

be about equal to the surface area of the projection lens so that light illuminating the slide is not wasted on the one hand, and on the other hand the full area of the projection lens is used for imaging the slide.

(d) Remember if two rays orignating from different parts of an object converge to a common point (after passing through some lens etc), that common point is *not* the image of anything. An image is where different rays from the *same* point of an object converge to. The image of the source is formed on the projection lens not on the slide.

(e) Because the image of the slide on the screen is real and inverted, the 'lower' portion of the image corresponds to 'upper' portion of the slide. But each point of the slide is illuminated by all points of the source. Thus every portion of the screen image gets illumination from all parts of the source.

11.78 (a) Image of a point object on a film (or a screen) is never strictly a point but is a small circle (called the circle of confusion) due to the lens aberration which can be reduced but never completely removed. Since aberrations reduce with decrease in aperture, the circle of confusion is smaller, the smaller the aperture of the lens. An object that is not properly focussed on the screen will naturally produce a large circle of confusion than the one that is properly focussed. Now there is a limit upto which our eye can perceive the size of a small circle. Below a certain angular size ($\sim 10^{-3}$ rad), our eye perceives a circle as a sharp point. Thus there is a *range of object distance* from the camera for which the images on the film have small enough circles of confusion for our eye to perceive them as sharp points. This range is called 'depth of field'. From what has been said above the depth of field will clearly increase if the aperture is reduced. For photographing a scenery, we will naturally like to have a greater depth of field than for an identity photograph.

(b) Field of view or the angle of view is determined by the size of the screen (i.e. the film) and the focal length of the camera lens (it is roughly (film size/focal length) in radians). To increase the field of view, one uses a lens of considerably smaller focal length than a normal camera lens (usually half of two thirds of the normal focal length of 5 cm). This lens is known as '*wide angle lens*'

(c) A telephoto lens is the opposite of wide-angle lens. It has a considerably larger focal length than a normal camera lens and therefore has a much reduced field view. Thus it photographs part of the object, but since image distance now

is more, that part looks more magnified. One feels that the object has been shot from a close distance.

(d) A telescope views large objects at large distances; a microscope views small objects at small distances. Both need a small field of view. A camera views objects of ordinary sizes at fairly close distances. Here the field of view required is much more (compare 45° for a camera with about 1° for a microscope objective and something similar for a telescope – a moon subtends about 0.5° at the earth). Thus rays entering a camera lens are far from being paraxial and aberrations will be large and images will be blurred if the apertures are not very small. For a telescope, on the other hand, the important thing is its ability to resolve distant objects (i.e., see them as distinct). We have seen the resolving power increases with increase in aperture. Therefore telescopes have as large an aperture as feasible.

11.79 (a) In a camera, image distance is nearly equal to the focal length of the lens (because $|u|/f \gg 1$). Therefore linear magnification is proportional to f. Therefore amount of light per unit area of the image is proportional to $1/f^2$. Amount of light is also proportional to exposure time t and to the square of aperture size: a^2. Thus the brightness of the image is proportional to a^2t/f^2. Therefore for a given brightness $t \propto f^2/a^2$. (Brightness is liminous flux per unit area.)

(b) The aperture size $f/5.6$ is about $\sqrt{2}$ times the aperture size $f/8$, i.e., the aperture $f/5.6$ gathers twice the amount of light per second gathered by the aperture $f/8$. To gather the same total amount of light, exposure time of $f/5.6$ should be half that of $f/8$ i.e., exposure time $1/120$ s.

11.80 The virtual image of the object (located at the focus of the objective) produced by L_1 should be located at the focus of L_2. Since $L_1\ L_2 = (2f/3)$, image distance from L_1, is: $v = -(f/3)$.

Using $$-\frac{1}{u} + \frac{1}{v} = \frac{1}{f},$$

$u = -(f/4)$. L_1 should be placed at a distance of $f/4$ from the focus of the objective.

Chapter 12

12.2 5.0×10^6 m s^{-1}

12.3 $a = 8.8 \times 10^{14}$ m s^{-2}, a antiparallel to **E**

12.4 8.8×10^{13} m s^{-2}

12.5 1.8×10^{11} C kg^{-1}

12.6 Use $\frac{e}{m} = \frac{E^2}{2B^2V}$ to get

$\frac{e}{m} = 1.8 \times 10^{11}$C kg^{-1}

12.8 4

12.9 $qE = \text{mg} = 6\pi\eta rv$;

$q'E = \text{mg} + 6\pi\eta rv = 2$ mg.

Thus $q' = 2q$; the drop has picked up 2 additional electrons from the surrounding air, i.e. in all, the drop now has 4 excess electrons.

12.12 1.5 eV $= 2.4 \times 10^{-19}$ J

12.13 6.59×10^{-34} Js (Standard value [upto three significant figures] $h = 6.63 \times 10^{-34}$ Js)

12.14 2.0V

12.15 No, because $\nu < \nu_o$

12.16 (a) True, (b) False, (c) True, (d) True.

12.17 1.53×10^{-10}m

12.18 1.2×10^{-12}m

12.19 1.2×10^{-10}m

12.21 6.634×10^{-21}J $= 4.15 \times 10^{-2}$ eV

12.22 $\lambda = \frac{h}{p} = \frac{h}{(h\nu/c)} = \frac{c}{\nu}$

12.23 α - particle

Answers to Additional Exercises

12.24 (a) Use $eV = (mv^2/2)$ i.e. $v = [(2eV/m)]^{1/2}$; $v = 1.33 \times 10^7$ ms^{-1}

(b) If we use the same formula with $V = 10^7$V, we get $v = 1.88 \times 10^9$ ms^{-1}. This is clearly wrong, since nothing can move with a speed greater than the speed of light ($c = 3 \times 10^8$ ms^{-1}). Actually the above formula for kinetic energy $[(mv^2/2)]$ is valid only when $(v/c) << 1$. At very high speeds when (v/c) is comparable to (though always less than) 1, we come to the relativistic domain where the following formulae are valid:

Relativistic momentum $p = mv$

Total energy $E = mc^2$

Kinetic energy,

$$T = mc^2 - m_0c^2,$$

where the relativistic mass m is given by

$$m = m_0\left(1 - \frac{v^2}{c^2}\right)^{-1/2};$$

m_0 is called the rest mass of the particle. These relations also imply

$$E = \left(p^2c^2 + m_0^2c^4\right)^{1/2}$$

Note that in the relativistic domain when v/c is comparable to 1, K.E. or energy $\geq m_0c^2$ (rest mass energy). The rest mass energy of electron is about 0.51 MeV. Thus a kinetic energy of 10 MeV, being much greater than electron's rest mass energy, implies relativistic domain. Using relativistic formulas v (for 10 MeV kinetic energy) = 0.999c.

12.25 (a) 22.7 cm.

(b) No. As explained above, a 20 MeV electron moves at relativistic speed. Consequently, the non-relativistic formula $R = (m_0v/eB)$ is not valid. The relativistic formula is

$$R = \frac{p}{eB} = \frac{mv}{eB}$$

$$\text{or} \quad R = \frac{m_0v}{eB\sqrt{1 - v^2/c^2}}$$

12.26 In this case $E = (V/d)$. Therefore $(e/m) = (V/2B^2d^2)$. Thus if V is doubled, B should be increased to $\sqrt{2}B$.

12.27 (a) The deflection at the other edge is given by

$$y = \frac{1}{2}\frac{eE}{m} \times \frac{L^2}{v^2} = \frac{eEL^2}{4eV}$$

$$= \frac{EL^2}{4V}$$

where E is the electric field between the plates and V is the accelerating voltage before the beam enters the plates. With the given rules, $y = 5.0$ mm.

(b) The slope of the parabolic trajectory at the edge is given, using the above relation between y and x (replace L by x):

$$\tan\theta = \left.\frac{dy}{dx}\right|_{x = L}$$

$$= \frac{EL}{2V} = \frac{EL^2}{4V}\Big/(L/2)$$

which shows that the tangent to the trajectory at the edge when produced back meets the initial direction of the beam at the mid-point $x = L/2$. The same tangent when produced further meets the screen at $y = Y$. Therefore, the deflection at the screen a distance D away from the far edge of the plates is given by :

$$Y = \left(D + \frac{L}{2}\right)\tan\theta$$

$$= \frac{EL}{2V}\left(D + \frac{L}{2}\right)$$

with the given values, $Y = 6.5$ cm.

12.28

$$Y = \frac{eEL}{2eV}\left(D + \frac{L}{2}\right)$$

$$= \frac{eEL}{mv^2}\left(D + \frac{L}{2}\right)$$

Using the 'no deflection' condition

$$v = \left|\frac{E}{B}\right|, \text{ this gives}$$

$$\frac{e}{m} = \frac{YE}{B^2L[D + (L/2)]}.$$

Using the given data,

$$\frac{e}{m} = 1.7 \times 10^{11} \text{ Ckg}^{-1}$$

12.29 We have $eV = (mv^2/2)$ and $R = (mv/eB)$ which gives $(e/m) = (2V/R^2B^2)$; using the given data $(e/m) = 1.73 \times 10^{11}$ Ckg^{-1}.

12.30 (a) Use $r^2 = \frac{9}{2}\frac{\eta.v}{(\rho - \sigma)g}$

to get $r = 7.26 \times 10^{-7}$ m ($\rho =$ density of drop, $\sigma=$ density of air)

(b) Use $qE = \frac{4\pi}{3}r^3(\rho - \sigma)g$

to get $q = 8.05 \times 10^{-19}$ C.

From the known value of electronic charge $e = 1.6 \times 10^{-19}$ C, it is clear that the drop carries 5 excess electrons.

12.31 (a) The radioactive source ionizes air through which the drop falls. The drop, therefore, picks up additional charges by acquiring excess electrons from the environment. It may also lose some of its electrons.

(b)
$$q_1E = \frac{4\pi}{3}r^3(\rho - \sigma)g - 6\pi\eta r v_1 = 6\pi\eta r(v_0 - v_1).$$

Charge quantization shows in the discrete values of $(v_0 - v)$ with $v = v_1, v_2, \cdots$ i.e. $(v_0 - v)$ are found to be integral multiples of a base value. (Note: if the net motion is upwards, v is negative).

12.32 Use the relations

$$mg = 6\pi\eta r v_1, \quad qE = 6\pi\eta r v_2$$

where v_1 and v_2 are the vertical and horizontal components of velocity of the drop. Therefore, if θ is the angle with the vertical,

$$\tan\theta = \frac{v_2}{v_1} = \frac{qE}{mg}$$

$$\text{or } q = \frac{4\pi}{3}\frac{r^3\rho g}{E}\tan\theta.$$

Using the given data, $q = 4.85 \times 10^{-19}$C. Clearly the drop has three excess electrons.

12.33 (a) 27.6 keV

(b) of the order of 30 kV

12.34 Use $\lambda = (hc/E)$ with $E = 5.1 \times 1.602 \times 10^{-10}$J to get $\lambda = 2.43 \times 10^{-16}$m.

12.35 (i) For $\lambda = 500$m, $E = (hc/\lambda) = 3.98 \times 10^{-28}$J. Number of photons emitted per second

$$= \frac{10^4 \text{J s}^{-1}}{3.98 \times 10^{-28}\text{J}} \simeq 3 \times 10^{31}\text{s}^{-1}$$

We see that the energy of a radiophoton is exceedingly small, and the number of photons emitted per second in a radio beam is enormously large. There is, therefore, negligible error involved in ignoring the existence of a minimum quantum of energy (photon) and treating the total energy of a radio wave as continuous.

(ii) For $\nu = 6 \times 10^{14}$ Hz, $E \simeq 4 \times 10^{-19}$J. Photon flux corresponding to minimum intensity

$$= \frac{10^{-10}\text{W}m^{-2}}{4 \times 10^{-19}\text{J}}$$

$= 2.5 \times 10^8 \mathrm{m}^{-2}\mathrm{s}^{-1}$.

Number of photons entering the pupil per second $= 2.5 \times 10^8 \times 0.4 \times 10^{-4}\ \mathrm{s}^{-1} = 10^4 \mathrm{s}^{-1}$.

Though this number is not as large as in (i) above, it is large enough for us never to 'sense' or 'count' individual photons by our eye.

12.36 $W_0 = h\nu - eV_0 = 6.7 \times 10^{-19}$ J

$$= 4.2\mathrm{eV};$$

$$\nu_0 = \frac{W_0}{h} = 1.0 \times 10^{15}\mathrm{Hz};$$

$$\lambda = 6328\ \text{Å}$$

corresponds to $\nu = 4.7 \times 10^{14}$ Hz $< \nu_0$. The photocell will not respond howsoever high be the intensity of laser light.

12.37 Use $eV_0 = h\nu - W_0$ for both sources. From the data on the first source, $W_0 = 1.40$eV. Use this value to obtain for the second source $V_0 = 1.50$V.

12.38 Obtain V_0 versus ν plot. The slope of the plot is h/e and its intercept on the ν-axis is ν_0. The first four points lie nearly on a straight line which intercepts the ν-axis at $\nu_0 = 5.0 \times 10^{14}$Hz. The fifth point corresponds to $\nu < \nu_0$; there is no p.e. emission and therefore no stopping voltage is required to stop the current. Slope of the plot is found to be 4.15×10^{-15}Vs. Using $e = 1.6 \times 10^{-19}$ C, $h = 6.64 \times 10^{-34}$Js (standard value $h = 6.626 \times 10^{-34}$ Js), $W_0 = h\nu_0 = 2.11$V.

12.39 It is found that the given incident frequency ν is greater than ν_0(Na), and ν_0(K); but less than ν_0(Mo), and ν_0(Ni). Therefore Mo and Ni will not give p.e. emission. If the laser is brought closer, intensity of radiation increases, but this does not affect the result regarding Mo and Ni; however p.e. current from Na and K will increase in proportion to intensity.

12.40 Assume one conduction electron per atom. Effective atomic area $\sim 10^{-20}\mathrm{m}^2$

Number of electrons in 5 layers

$$= \frac{5 \times 2 \times 10^{-4}\mathrm{m}^2}{10^{-20}\mathrm{m}^2}$$
$$= 10^{17}$$

Incident power

$$= 10^{-5}\mathrm{Wm}^{-2} \times 2 \times 10^{-4}\mathrm{m}^2$$
$$= 2 \times 10^{-9}\mathrm{W}$$

In the wave picture, incident power is uniformly absorbed by all the electrons continuously. Consequently, Energy absorbed per second per electron

$$= \frac{2 \times 10^{-9}}{10^{17}} = 2 \times 10^{-26}\mathrm{W}.$$

Time required for p.e. emission

$$= \frac{2 \times 1.6 \times 10^{-19}\mathrm{J}}{2 \times 10^{-26}\mathrm{W}} = 1.6 \times 10^7\mathrm{s}$$

which is about 0.5 year.

Implication: Experimentally, p.e. emission is observed nearly instantaneously ($\sim 10^{-9}$s): Thus the wave picture is in gross disagreement with experiment. In the photon picture, energy of the radia-

tion is not continuously shared by all the electrons in the top layers. Rather, energy comes in discontinuous 'quanta'. And absorption of energy does not take place gradually. A photon is either not absorbed, or absorbed by an electron nearly instantly.

12.41 (a) The conditions of energy and momentum conservations are:

$$m_0c^2 + h\nu = \sqrt{m_0^2c^4 + p^2c^2}$$

$$0 + \frac{h\nu}{c} = p$$

These conditions imply

$$2m_0c^3p = 0,$$

which is impossible. In a frame where the initial electron is moving with uniform velocity, the same conclusion must hold because if a process is forbidden in one inertial frame, it is also forbidden in another inertial frame.

(b) We have shown in (a) that

$$e^- + \gamma \rightarrow e^-$$

(forbidden). However, for an electron in a lattice, the momentum of the incident photon can be shared by both the electron and the lattice, while the lattice due to its very large mass (compared to the mass of the electron) does not share the energy of the incident photon. This situation is like a ball rebounding from a wall, where the wall shares momentum but not energy. See chapter 6 (Class XI). In short, while

$$e^- + \gamma \rightarrow e^- \quad \text{is forbidden,}$$

$$e^- + \gamma + \text{lattice} \rightarrow e^- + \text{lattice}$$

is not forbidden.

12.42 Use $h\nu - B = E$; B = binding energy of the level from which electron is emitted after absorption of photon. E is the energy of the photoelectron emitted. For a given ν, E is discrete because B is discrete. Now $B + E$ = 347 keV (K shell), 352 keV (L shell), 357.5 keV (M shell). The values of $B + E$ are equal within about 3%. Taking $B + E \simeq 350$ keV, the wavelength for γ-ray is estimated to be $\lambda = hc/(B + E) = 3.5 \times 10^{-12}$m.

12.43 Use $R = \alpha E$ to obtain

$$E_1 = \frac{R_1}{\alpha} = \frac{1.40\text{cm}}{1\text{cm keV}^{-1}} = 1.40\text{keV}$$

Similarly $E_2 = 2.02$ keV.

Incident photon energy $= (hc/\lambda) =$ 2.53 keV

Use $h\nu - B = E$ to get

$$B_1 = 1.13\,\text{keV}, \quad B_2 = 0.51\,\text{keV}$$

12.44 For λ = 1Å, electron's energy = 150 eV; photon's energy = 12.4 keV. Thus for the same wavelength, a photon has much greater energy than an electron.

12.45 (a) $\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m \times K.E}}$

Thus for the same $K.E.$, λ decreases with m as $(1/\sqrt{m})$. Now $(m_n/m_e) = 1838.6$; therefore for the same energy, (150 eV) as in 12.44, wavelength

of neutron $= (1/\sqrt{1838.6}) \times 10^{-10}$m $= 2.33 \times 10^{-12}$m. The interatomic spacing is about a hundred times greater. A neutron beam of 150 eV energy is therefore not suitable for diffraction experiments.

(b) $\lambda = 1.45 \times 10^{-10}$m (Use $\lambda = (h/\sqrt{3mkT})$) which is comparable to interatomic spacing in a crystal.

Clearly from (a) and (b) above, thermal neutrons are a suitable probe for diffraction experiments; so a high energy neutron beam should be first thermalized before using it for diffraction.

12.46 $\lambda = 5.5 \times 10^{-12}$m

λ (yellow light) $= 5.9 \times 10^{-7}$m.

RP is inversely proportional to wavelength. Thus RP of an electron microscope is about 10^5 times that of an optical microscope. In practice differences in other (geometrical) factors can change this comparison somewhat.

12.47 Use

$$p = \frac{h}{\lambda} = \frac{6.63 \times 10^{-34}\text{Js}}{10^{-15}\text{m}} = 6.63 \times 10^{-19}\text{kg ms}^{-1}$$

Use the relativistic formula for energy:

$$\begin{aligned} E^2 &= c^2p^2 + m_0^2c^4 \\ &= 9 \times (6.63)^2 \times 10^{-22} \\ &\quad +(0.511 \times 1.6)^2 \times 10^{-26} \\ &\simeq 9 \times (6.63)^2 \times 10^{-22}, \end{aligned}$$

the second term (rest mass energy) being negligible. Therefore $E = 1.989 \times 10^{-10}$ J $= 1.24$ BeV. Thus electron energies from the accelerator must have been of the order of a few BeV.

12.48 For $\lambda = 10^{-14}$m, E=124 MeV. this is much too large compared to the binding energy that Coulomb force can provide within the nucleus. Therefore, electrons localized within a nucleus are far too energetic to stay bound within. This is why electrons do not reside in a nucleus.

12.49 Use

$$\lambda = \frac{h}{\sqrt{3mkT}};$$

$$m_{\text{He}} = \frac{4 \times 10^{-3}}{6 \times 10^{23}}\text{kg}.$$

This gives $\lambda = 0.73 \times 10^{-10}$m.

Mean separation

$$r = \left(\frac{V}{N}\right)^{1/3} = \left(\frac{kT}{p}\right)^{1/3}$$

For $T = 300$K, $p = 1.01 \times 10^5$ Pa , $r = 3.4 \times 10^{-9}$m. We find $r >> \lambda$.

12.50 Using the same formula as in 12.49, $\lambda = 6.2 \times 10^{-9}$m which is much greater than the given inter-electron separation.

12.51 (a) Quarks are thought to be confined within a proton or neutron by forces which grow stronger if one tries to pull them apart. It, therefore, seems that though fractional charges may exist in nature,

observable charges are still integral multiples of e.

(b) Electric fields needed in the experiment will be impractically high.

(c) Stokes' formula is valid for motion through a homogeneous continuous medium. The size of the drop should be much larger than the intermolecular separation in the medium for this assumption to be valid; otherwise the drop 'sees' inhomogeneities in the medium (concentrated mass density in molecules, and holes in between molecules)

(d) Both the basic relations

$$eV = \frac{1}{2}mv^2$$
$$\text{or} \quad eE = ma$$
$$\text{and} \quad eBv = \frac{mv^2}{r}$$

for electric and magnetic fields respectively show that the dynamics of electrons is determined not by e, and m separately but by the combination e/m.

(e) At low pressures, ions have a chance to reach their respective electrodes and constitute a current. At ordinary pressures, ions have no chance to do so because of collisions with gas molecules and recombination.

(f) Work function merely indicates the minimum energy required for the electron in the highest level of the conduction band to get out of the metal. Not all electrons in the metal belong to this level. They occupy a continuous band of levels. Consequently, for the same incident radiation, electrons knocked off from different levels come out with different energies.

(g) The absolute value of energy E (but not momentum p) of any particle is arbitrary to within an additive constant. Hence while λ is physically significant, absolute value of ν of a matter wave of an electron has no direct physical meaning. The phase speed $\nu\lambda$ is likewise not physically significant. The group speed given by

$$\frac{\mathrm{d}\nu}{\mathrm{d}(1/\lambda)} = \frac{\mathrm{d}E}{\mathrm{d}p} = \frac{\mathrm{d}}{\mathrm{d}p}\left(\frac{p^2}{2m}\right) = \frac{p}{m}$$

is physically meaningful.

Chapter 13

13.1 2.3×10^{-14} m

13.2 1.8×10^{-14} m

13.3 5.6×10^{14} Hz

13.6 13.6 eV; – 27.2 eV

13.7 1217 Å

13.9 2.12×10^{-10}m ; 4.77×10^{-10} m

13.10 104.7 MeV

13.12 (i) ${}^{226}_{88}\text{Ra} \longrightarrow {}^{222}_{86}\text{Rn} + {}^{4}_{2}\text{He}$

(ii) ${}^{32}_{15}\text{P} \longrightarrow {}^{32}_{16}\text{S} + \bar{e} + \bar{\nu}$

(iii) ${}^{11}_{6}\text{C} \longrightarrow {}^{11}_{5}\text{B} + e^{+} + \nu$

13.13 4 T years

13.14 7.88×10^{10} s^{-1}

13.15 7.1 mg

13.16–6.10 eV

Answers to Additional Exercises

13.18 (a) No different from

(b) Thomson's model ... Rutherford's model.

(c) Rutherford's model

(d) Thomson's model ... Rutherford's model

(e) both the models

13.19 (a) about the same

(b) much less

(c) It suggests that scattering is predominantly due to a single collision, because the chance of a single collision increases linearly with the number of target atoms, and hence linearly with the thickness of the foil.

(d) In Thomson's model, a single collision causes very little deflection. The observed average scattering angle can be explained only by considering multiple scattering. So it is wrong to ignore multiple scattering in Thomson's model. In Rutherford's model, most of the scattering comes through a single collision and multiple scattering effects can be ignored as a first approximation.

13.20 Angular momentum conservation gives $mvb = mv's$ (*Note*: at the point of minimum distance, velocity is normal to the radius vector from the nucleus to the point) Energy conservation gives

$$\frac{1}{2}mv^2 = \frac{1}{2}mv'^2 + \frac{ZZ'e^2}{4\pi\epsilon_0 s}$$

Eliminate v' to get

$$s^2 = \frac{ZZ'e^2 s}{2\pi\epsilon_0 mv^2} + b^2$$

For

$$b = 0, s = r_0 = \frac{ZZ'e^2}{4\pi\epsilon_0\,(mv^2/2)}$$

13.21 Thomson's model predicts

$$P(\geq 90°) = e^{-(90/1)^2}$$

i.e. $\log_{10} P = -3518$

in gross disagreement with the observed result: $\log_{10} P = -4$

13.22 (a) $b = 0$ implies $\cot(\theta/2) = 0$ i.e. $\theta/2 = 90°$ or $\theta = 180°$, as expected physically,

(b) For a given b, increase in energy implies increase in $\cot(\theta/2)$ and hence decrease in scattering angle, as intuitively expected.

(c) $b = 1.1 \times 10^{-14}$m

(d) Charge of the nucleus is what provides the field due to which scattering takes place. If $Z = 0$, $\theta = 0$ from the formula, as expected. Mass of the nucleus does not enter since recoil of the nucleus is being ignored. If recoil is included, a small correction term to the formula will include mass of the nucleus.

(e) For a given energy, decrease in b implies decrease in $\cot(\theta/2)$ and hence increase in scattering angle, as expected physically.

13.23 The first orbit in Bohr's model has a radius a_0 given by $a_0 = (4\pi\epsilon_0\hbar^2)/(m_e e^2)$. If in place of the electrostatic force $e^2/(4\pi\epsilon_0 r^2)$ we consider the atom bound by the gravitational force $(Gm_p m_e/r^2)$ we should replace $(e^2/4\pi\epsilon_0)$ by $Gm_p m_e$. That is, the radius of the first Bohr's orbit in a gravitationally bound hydrogen atom is given by $a_0^G = (\hbar^2/Gm_p m_e^2) \simeq 1.2 \times 10^{29}$m. This is much greater than the estimated size of the whole universe!

13.24 (i) The energy required to excite the electron from the ground state ($n = 1$) to the $n = 3$ state is $E_3 - E_1$

$$= \frac{mZ^2e^4}{32\pi^2\epsilon_0^2\hbar^2}\left[\frac{1}{1^2} - \frac{1}{3^2}\right]$$
$$= 48.4\text{eV}$$

where $Z = 2$.

(ii) Ionization energy is given by

$$E_\infty - E_1$$
$$= \frac{mZ^2e^4}{32\pi^2\epsilon_0^2\hbar^2}(\text{with } Z = 3)$$
$$\simeq 122\text{eV}.$$

Ionization potential is 122 V.

13.25

$$\nu = \frac{me^4}{(4\pi)^3\epsilon_0^2\hbar^3}\left[\frac{1}{(n-1)^2} - \frac{1}{n^2}\right] = \frac{me^4(2n-1)}{(4\pi)^3\epsilon_0^2\hbar^3 n^2(n-1)^2}$$

For large $n, \nu \simeq (me^4/32\pi^3\epsilon_0^2\hbar^3 n^3)$ Orbital frequency $\nu_c = (v/2\pi r)$. Now in Bohr's model, $v = (n\hbar/mr)$ and $r = (4\pi\epsilon_0\hbar^2/me^2)n^2$. This gives $\nu_c = (n\hbar/2\pi m r^2) = (me^4/32\pi^3\epsilon_0^2\hbar^3 n^3)$ same as ν above for large n.

13.26 $r_n = \dfrac{4\pi\epsilon_0\hbar^2}{Ze^2m}n^2$ i.e. $r_n \propto \frac{n^2}{Z}$.

For hydrogen $Z = 1$, for Be^{+++}, $Z = 4$. The ($n = 2$) state of Be^{+++} has the same radius as ($n = 1$) state of hydrogen. Now $E_n \propto Z^2/n^2$. Therefore $(E_2(Be^{+++})/E_1(\text{H}))=4$.

13.27 $E_n \propto Z^2/n^2$ for Li^{++}, $Z = 3$. Therefore (n=3) state of Li^{++} has the

same energy as (n=1) state of hydrogen. Now $r_n \propto (n^2/Z)$. Therefore $r_3(Li^{++}) = 3r_1(\mathrm{H})$.

13.28 In Bohr's model, $mvr = n\hbar$ and

$$\frac{mv^2}{r} = \frac{Ze^2}{4\pi\epsilon_0 r^2}$$

which give

$$T = \frac{1}{2}mv^2 = \frac{Ze^2}{8\pi\epsilon_0 r};$$

$$r = \frac{4\pi\epsilon_0\hbar^2}{Ze^2 m}n^2$$

These relations have nothing to do with the choice of the zero of potential energy. Now, choosing the zero of potential energy at infinity we have $V = -(Ze^2/4\pi\epsilon_0 r)$ which gives $V = -2T$ and $E = T + V = -T$

(a) The quoted value of $E = -3.4$ eV is based on the customary choice of zero of potential energy at infinity. Using $E = -T$, the kinetic energy of the electron in this state is $+3.4$ eV.

(b) Using $V = -2T$, potential energy of the electron is $= -6.8$ eV

(c) If the zero of potential energy is chosen differently, kinetic energy does not change. Its value is $+3.4$ eV independent of the choice of the zero of potential energy. The potential energy, and the total energy of the state, however, would alter if a different zero of the potential energy is chosen.

13.29 $v = \dfrac{n\hbar}{mr}$, $r = \dfrac{4\pi\epsilon_0 n\hbar}{Ze^2 m}n^2$

That is $v = \dfrac{Ze^2}{4\pi\epsilon_0 n\hbar}$.

For $n = 3, Z = 2$,

$v = 1.46 \times 10^6$ ms^{-1}.

Thus $(v/c) \simeq 0.005$; non-relativistic approximation is valid because $(v/c) << 1$.

13.30 Angular momenta associated with planetary motion are incomparably large relative to $\hbar$. For example, angular momentum of the earth in its orbital motion is of the order of $10^{70}\hbar$. In terms of the Bohr's quantization postulate this corresponds to a very large value of n (of the order of 10^{70}). For such large values of n, the differences in the successive energies and angular momenta of the quantized levels of the Bohr model are so small compared to the energies and angular momenta respectively for the levels that one can, for all practical purposes, consider the levels continuous.

13.31 (a) The quantity $(e^2/4\pi\epsilon_0 mc^2)$ has the dimension of length. Its value is 2.82×10^{-15} m - much smaller than the typical atomic size.

(b) The quantity $4\pi\epsilon_0\hbar^2/me^2$ has the dimension of length. Its value is 0.53×10^{-10}m - of the order of atomic sizes. (Note: dimensional arguments cannot, of course, tell us that we should use 4π factors and $\hbar$ in place of h to arrive at the right size).

13.32 Ground state energy of a hydro-

gen atom = −13.6 eV. Energy of each electron in helium ground state $= -Z_{eff}^2 \times 13.6$ eV. Helium ground state energy $= -Z_{eff}^2 \times 27.2$ eV.

Energy of He$^+$ ground state $= -4 \times 13.6 = -54.4$ eV (Note in He$^+$ ground state, Z=2; there is no screening because there is just one electron).

Ionization potential $= [-54.4 + Z_{eff}^2 \times 27.2]$ V = 24.46 V (experimental value) which gives $Z_{eff} =$ 1.70

13.33 All that is needed is to replace m_e by m_μ in the formulas of the Bohr model. We note that keeping other factors fixed, $r \propto (1/m)$ and $E \propto m$.

Therefore

$$r_\mu = \frac{r_e m_e}{m_\mu} = \frac{0.53 \times 10^{-13}}{207}$$
$$= 2.56 \times 10^{-13}\text{m}$$
$$E_\mu = \frac{E_e m_\mu}{m_e} = -(13.6 \times 207)\text{eV}$$
$$\simeq -2.8\text{keV}$$

13.34 Radius of the ($n = 1$) orbit of the π^- -mesic atom (ground state)

$$r_{\pi^-} = \frac{4\pi\epsilon_0\hbar^2}{m_\pi e^2},$$

ignoring the correction due to finite mass of the proton $r_{\pi^-} = 1.94 \times 10^{-13}$m. Next, the orbital speed is given by

$$v = \frac{e^2}{4\pi\epsilon_0\hbar} = 2.19 \times 10^6 \text{ ms}^{-1}.$$

Time required for one revolution,

$$T = \frac{2\pi r_{\pi^-}}{v} = 5.6 \times 10^{-19}\text{s}$$

Estimated number of revolutions before decay

$$\approx \frac{10^{-8}}{5.6 \times 10^{-19}} \approx 2 \times 10^{10}.$$

13.35 In an ordinary atom, motion of the nucleus can be ignored (as a first approximation). In a positronium atom, one must considerd the motion of both electron and positron about their centre of mass. A detailed analysis (beyond our scope here) shows that formulas of the Bohr model apply provided we replace m_e by what is known as the 'reduced mass' of the electron. For an e^-e^+ system, the reduced mass is $m_e/2$. From the answer to 13.7, we then get the wavelength of radiation emitted in the n=2 to n=1 transition to be 2 × 1217 Å= 2434 Å which is in the ultraviolet part of the e.m. spectrum.

13.36 Nuclear mass density is of the order of 10^{17} kgm^{-3}. This estimate is obtained by using the empirical relation: $r = r_0 A^{1/3} (r_0 = 1.1 \times 10^{-15}\text{m})$ for the radius of a nucleus of mass number A. It is typically 10^{13} to 10^{14} times the average mass density of an atom. That is, the average atomic mass density is of the order of 10^3 to 10^4 kgm^{-3}. Typical mass densities of solids and liquids are of the same order since atoms are tightly packed in these phases. Typical densities of gases at STP are of the order of 10^{-1} to 1 kg m^{-3}.

13.37 $\dfrac{r(^{197}\text{Au})}{r(^{107}\text{Au})} = \left(\dfrac{197}{107}\right)^{1/3} \simeq 1.23$

The ratio of their nuclear mass densities is roughly one.

13.38 The average atomic mass of neon is

$$m((\text{Ne})) = [90.51 \times 19.99 + 0.27 \times 20.99 + 9.22 \times 21.99] \times 10^{-2}\ \text{u} = 20.18\ \text{u}$$

13.39 (a) There are several reasons as to why the atomic mass of an isotope of mass number A is *not exactly* A times the atomic mass of a hydrogen atom. First, the proton and neutron masses are not identical. Second, the atomic mass also includes the mass of electrons and an atom of mass number A does not have A but has Z electrons. Third, the mass of a nucleus is slightly less than the sum of the masses of its consitituent protons and neutrons. This mass defect Δm is related to the binding energy of the nucleus via the Einstein's relation $\Delta m(c)^2 = B.E.$ Mass defect arises also due to the binding energy of the atom (i.e. electronic binding energy) but this is so small compared to the effect of the above three factors that it can be ignored.

(b) The ratio of $m(^{16}_{8}\text{O})$ and $m(^{8}_{4}\text{Be})$ is not exactly 2 because of the third reason mentioned in (a) above. The precise value of binding energy depends on the details of nuclear dynamics. There is *no* reason why the binding energy of $(^{16}_{8}\text{O})$ should be *precisely* twice the binding energy $(^{8}_{4}\text{Be})$. This is what causes departure from the value 2 for the ratio of their masses.

13.40 Binding energy of a nucleus $^{A}_{Z}X$ is the energy needed to split the nucleus into Z protons and $(A-Z)$ neutrons. Using mass-energy equivalence,

$$m_N(^{A}_{Z}X)c^2 + B.E. = [Zm_pc^2 + (A-Z)m_nc^2]$$

where it is important to note that m_N represents nuclear not atomic mass of $^{A}_{Z}X$. That is

$$B.E. = [Zm_p + (A-Z)m_n - m_N(^{A}_{Z}X)]\,c^2.$$

By the same reasoning, atomic mass $m(^{A}_{Z}X)$ is related to $m_N(^{A}_{Z}X)$ by the relation

$$m(^{A}_{Z}X) = m_N(^{A}_{Z}X) + Zm_e - \frac{1}{c^2} \times \begin{pmatrix}\text{binding energy}\\ \text{of } Z \text{ electrons}\end{pmatrix}.$$

Similarly

$$m_H = m_p + m_e - \frac{1}{c^2} \times \begin{pmatrix}\text{binding energy of}\\ \text{the electron in H}\\ \text{atom}\end{pmatrix}.$$

Replacing m_p in terms of m_H, and $m_N(^{A}_{Z}X)$ in terms of $(^{A}_{Z}X)$ we note that the Zm_e terms cancel away. Further, electron binding energies (of the order of eV to keV) are much smaller than nuclear binding energies. It is, therefore, a safe approximation to ignore terms containing electronic binding energies. We then get the required formula for B.E.

of a nucleus in terms of its atomic mass.

13.41 $1\text{u} = 1.660565 \times 10^{-27}$ kg.
$1\text{u} \times c^2 \simeq 931.5$ MeV.

Using the formula for B.E. given in 13.40 we get
B.E. $({}^{56}_{26}\text{Fe}) = 492.26$ MeV
B.E. per nucleon $\doteq 8.790$ MeV.
B.E. $({}^{209}_{83}\text{Bi}) = 1640.30$ MeV
B.E. per nucleon $= 7.85$ MeV
^{56}Fe has greater binding energy per nucleon.

13.42 Neutron separation energy S_n of a nucleus ${}^{A}_{Z}X$ is given by

$$S_n = [m_N({}^{A-1}_{Z}X) + m_n - m_N({}^{A}_{Z}X)]c^2$$

Adding and subtracting the term Zm_e in the bracket above and ignoring mass defects due to electronic binding energies, we get S_n in terms of atomic masses.

$$S_n = [m({}^{A-1}_{Z}X) + m_n - m({}^{A}_{Z}X)]c^2.$$

From the given data, using $c^2 =$ 931.5 MeV/u we get
$S_n({}^{41}_{20}\text{Ca}) = 8.36$ MeV
$S_n({}^{27}_{13}\text{Al}) = 13.06$ MeV

Data on neutron separation energies provide important hints regarding the shell structure of a nucleus, as you will learn in more advanced courses.

13.43 ${}^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\text{He} + (Q)$ where Q (called Q-value) represents kinetic energy released. From Einstein's mass-energy equivalence

$$Q = [m_N({}^{238}\text{U}) - m_N({}^{234}\text{Th}) - m_N({}^{4}\text{He})]c^2$$

Add and subtract $92m_e$ in the bracket to write Q in terms of atomic masses (ignoring mass defects due to electronic binding energies as before):

$$Q = [m({}^{238}\text{U}) - m({}^{234}\text{Th}) - m({}^{4}\text{He})]c^2$$

Using the given data and $c^2 = 931.5$ MeV/u, Q=4.26 MeV. This total kinetic energy is shared by the ^{234}Th nucleus and the α-particle. In the rest frame of ^{238}U, ^{234}Th and the α-particle have equal and opposite momenta. Consequently, since ^{234}Th is much more massive than ^{4}He, the kinetic energy of ^{234}Th is much less than that of ^{4}He. (recall the formula : kinetic energy $= p^2/2M$). Thus most of the kinetic energy released (= 4.27 MeV) is carried by the α -particle.

13.44 (a) The rate of a nuclear reaction (which is energetically allowed) depends on the dynamical details of the reaction. α -decay is caused by the quantum mechanical tunneling of an α -particle through a repulsive Coulomb barrier. The rate of this tunneling depends on the details of this barrier such as its height, width etc.

(b) Simple. Because it is not energetically allowed. You can verify from the data that $m({}^{237}_{92}\text{U}) + m_n > m({}^{238}_{92}\text{U})$. This means spontaneous emission of neutron is not possible. Rather, one would need to suply energy to separate a neutron from ${}^{238}_{92}$U.

13.45 $^{23}_{10}\text{Ne} \longrightarrow ^{23}_{11}\text{Na} + e^- + \bar{\nu}_e + (Q)$

$$Q = [m_N(^{23}_{10}\text{Ne}) - m_N(^{23}_{11}\text{Na}) - m_e]c^2$$

where the neutrino rest mass has been ignored. Thus

$$\begin{aligned} Q &= \left[m(^{23}_{10}\text{Ne}) - 10m_e - m(^{23}_{11}\text{Na}) \right. \\ &\quad \left. + 11m_e - m_e\right] c^2 \\ &= \left[m(^{23}_{10}\text{Ne}) - m(^{23}_{11}\text{Na})\right] c^2 \\ &= 4.374\,\text{MeV}. \end{aligned}$$

This kinetic energy is mainly shared jointly by the $e^- - \bar{\nu}_e$ pair since ^{23}Na is much more massive than this pair and its recoil energy is therefore negligible. The maximum energy of emission is equal to the total kinetic energy of the $e^- - \bar{\nu}_e$ pair. (When the electron has maximum kinetic energy, the neutrino carries no energy). Thus the maximum kinetic energy of the β^- emitted is 4.374 MeV.

13.46 (a) Consider the decay of a free neutron at rest:

$$n \rightarrow p + e^-$$

If the momentum of e^- is $\bar{P}_e$, momentum of the proton is $-\bar{P}_e$ by momentum conservation. Next, by total energy conservation,

$$\sqrt{c^2P_e^2 + m_p^2c^4} + \sqrt{c^2P_e^2 + m_e^2c^4} = m_nc^2.$$

Thus there is a definite momentum P_e given by the above equation in terms of constants. This means the energy of the electron in the above decay is fixed and determined in terms of the masses of the particles involved. It is thus impossible for the electron in the above decay to have a continuous distribution of energies. The presence of an additional particle, however, allows this possibility. The available energy can now be shared by the electron and the third particle, and the electron's energy is no longer fixed. This simple reasoning was among the several arguments that led Pauli to explain the observed continuous distribution of electron energy in β-decay by postulating the existence of a new particle till then unobserved. We now know that the correct equation for β-decay is : $n \rightarrow p + e^- + \bar{\nu}_e$ where the particle denoted by $\bar{\nu}_e$ is called the (electron) antineutrino. It is a neutral particle of negligibly small rest mass and intrinsic spin 1/2. It is extremely difficult to detect a neutrino because of its weak interaction with matter.

(b) A free neutron has rest mass greater than that of a proton plus electron. Thus β^-–decay is energetically allowed, but the β^+ -decay of a free proton into a neutron is not allowed: In a nucleus, individual neutrons and protons are not free. Thus the β^+–decay of a proton ($p \rightarrow n + e^+ + \nu_e$) *is possible when the proton is bound in a nucleus*. The energy needed for the decay can come from the appropriate difference in binding energies of a

proton and a neutron in the nucleus. In a stable nucleus with Z protons and $(A-Z)$ neutrons, the two reciprocal processes (neutron decay and proton decay) are in dynamic equilibrium.

13.47 Consider the two competing processes:

$${}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + e^{+} + \nu_e + Q_1$$
(positron emission)

$$e^{-} + {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + \nu_e + Q_2$$
(electron capture)

$$\begin{aligned} Q_1 &= [m_N({}^{A}_{Z}X) - m_N({}^{A}_{Z-1}Y) \\ &\quad - m_e]c^2 \\ &= [m({}^{A}_{Z}X) - Zm_e - m({}^{A}_{Z-1}Y) + \\ &\quad (Z-1)m_e - m_e]c^2 \\ &= [m({}^{A}_{Z}X) - m({}^{A}_{Z-1}Y) - 2m_e]c^2 \end{aligned}$$

$$\begin{aligned} Q_2 &= [m_N({}^{A}_{Z}X) + m_e - \\ &\quad m_N({}^{A}_{Z-1}Y)]c^2 \\ &= [m({}^{A}_{Z}X) - m({}^{A}_{Z-1}Y)]c^2 \end{aligned}$$

This means $Q_1 > 0$ implies $Q_2 > 0$ but $Q_2 > 0$ does not necessarily mean $Q_1 > 0$. Hence the result.

13.48
$$\begin{aligned} \nu(\gamma_1) &= 2.608 \times 10^{20}\text{Hz} \\ \nu(\gamma_2) &= 9.950 \times 10^{19}\text{Hz} \\ \nu(\gamma_3) &= 1.633 \times 10^{20}\text{Hz} \end{aligned}$$

These frequencies are obtained by dividing energy differences by h. Maximum kinetic energy of the β-particles:

$K_{max}(\beta_1^-) =$
$c^2[m({}^{198}_{79}\text{Au})$ – mass of second excited state of ${}^{198}_{80}\text{Hg}]$

$$= c^2 \left[m({}^{198}_{79}\text{Au}) - \left\{ m({}^{198}_{80}\text{Hg}) + \frac{1.088}{931.5} \right\} \right]$$

Use $c^2 = 931.5$ MeV/u to get

$$\begin{aligned} K_{max}(\beta_1^-) &= [931.5\{m({}^{198}_{79}\text{Au}) \\ &\quad - m({}^{198}_{80}\text{Hg})\} - 1.088] \\ &= 0.281 \text{ MeV}. \end{aligned}$$

Similarly $K_{max}(\beta_2^-) = 0.957$ MeV.

13.49 $N = N_0 e^{-\lambda t}$. Now activity is proportional to the number of radioactive atoms, so $e^{-\lambda t} = (9/15)$ i.e. $t = (1/\lambda)\log_e(5/3)$
λ is related to half-life $t_{1/2}$ by $\lambda = (\log_e 2/t_{1/2})$
Thus

$$\begin{aligned} t &= \frac{\log_e(5/3)}{\log_e 2} \times t_{1/2} \\ &= \frac{\log_{10}(5/3)}{\log_{10}(2)} \times t_{1/2} \\ &= 4224 \text{ years}. \end{aligned}$$

13.50 (i) $Q = -4.03$ MeV; reaction is endothermic.

(ii) $Q = +4.62$ MeV; reaction is exothermic.

13.51 (a) A chemical equation is balanced in the sense that the number of atoms of each element is the same on both sides of the equation. A chemical reaction merely alters the original combinations of atoms. In a nuclear reaction, elements may be transmuted. Thus the number of atoms of

each element is not necessarily conserved in a nuclear reaction. However, the number of protons and the number of neutrons are both separately conserved in a nuclear reaction. [Actually even this is not strictly true in the realm of very high energies — what is strictly conserved is the total charge and total 'baryon number'. We need not pursue this matter here]. In the reactions of 13.50, the number of protons and the number of neutrons are the same on the two sides of each equation.

(b) We know that the binding energy of a nucleus gives a negative contribution to the mass of the nucleus (mass defect). Now since proton number and neutron number are conserved in a nuclear reaction, the total rest mass of neutrons and protons is the same on either side of a reaction. But the total binding energy of nuclei on the left side need not be the same as that on the right hand side. The difference in these binding energies appears as energy released or absorbed in a nuclear reaction. Since binding energy contributes to mass, we say that the difference in the total mass of nuclei on the two sides gets converted into energy or vice versa. It is in this sense that a nuclear reaction is an example of mass-energy interconversion.

(c) From the point of view of mass energy interconversion, a chemical reaction is similar to a nuclear reaction *in principle*. The energy released or absorbed in a chemical reaction can be traced to the difference in chemical (not nuclear) binding energies of atoms and molecules on the two sides of a reaction. Since, strictly speaking, chemical binding energy also gives a negative contribution (mass defect) to the total mass of an atom or molecule, we can equally well say that the difference in the total mass of atoms or molecules on the two sides of a chemical reaction gets converted into energy or vice versa. However, the mass defects involved in a chemical reaction are almost a million times smaller than those in a nuclear reaction. This is the reason for the general impression, (which is *incorrect*) that mass-energy interconversion does not take place in a chemical reaction.

13.52 The net fission process is

$$^{235}_{92}\mathrm{U} + n \rightarrow {}^{94}_{40}\mathrm{Zr} + {}^{140}_{58}\mathrm{Ce} + 2n + 6e^- + (Q)$$

$$Q = [m_N(^{235}\mathrm{U}) - m_N(^{94}\mathrm{Zr}) - m_N(^{140}\mathrm{Ce}) - m_n - 6m_e]c^2$$

where we have ignored the very small energy of thermal neutrons. Converting the nuclear masses into atomic masses,

$$Q = [m(^{235}\mathrm{U}) - 92m_e - m(^{94}\mathrm{Zr}) + 40m_e - m_e(^{140}\mathrm{C}) + 58m_e$$

$$- m_n - 6m_e]c^2$$
$$= [m(^{235}\text{U}) - m(^{94}\text{Zr}) - m(^{140}\text{Ce}) - m_n]c^2 .$$

Using the given data $Q = 208$ MeV.

All this energy is not available for fission products only. Part of it (about 20%) is carried by the neutrons produced in the fission process as also by the β -particles produced in the radioactive decay of the initial fission products.

13.53 Required power from nuclear plants $= 10^{10}$W
Required electric energy from nuclear plants in one year $=3.156 \times 10^{17}$J
Required number of fissions per year $= 3.945 \times 10^{28}$
Available electric energy per fission $= 0.25 \times 200 = 50$ MeV $= 8 \times 10^{-12}$J
Required number of moles of ^{235}U $= 6.55 \times 10^4$
Required mass of ^{235}U $= 6.55 \times 235 \times 10^4$g $= 1.54 \times 10^4$ kg.

13.54 (a) (i) The neutrons produced in fission have energies of the order of a few MeV. The cross-section (i.e. probability) for fission of ^{235}U by these fast neutrons is negligible compared to that by slow (thermal) neutrons. The role of the moderator is to slow down neutrons for further fission of ^{235}U and thus sustain a chain reaction. Now energy loss of a neutron in a collision is maximum if it hits a nucleus of the same mass. Thus ordinary water (consisting of hydrogen atoms) can be used as an effective moderator. There is, however, one problem. In a neutron - proton collision, there is a considerable chance of another process, namely absorption of a neutron by a proton via the reaction. $n + p \rightarrow d + \gamma$ (d; deuteron). To avoid this difficulty, one uses heavy water as a moderator, which has negligible cross-section for neutron absorption.

(ii) For a controlled chain reaction, the average number of available neutrons should never exceed one per fission. Any excess neutrons over this 'critical limit' should be absorbed. This is what the control rods do. They are made of cadmium because cadmium has a high cross-section for neutron absorption.

(iii) Some (about 0.01 of the total) of the neutrons produced in fission are delayed by a few seconds because they are produced in subsequent decays of the initial fission fragments. This circumstance is crucial to mechanical control of the reactor. If all the fission neutrons were produced instantly in fis-

sion, there would be no time for the minute adjustments required in a reactor to keep it critical.

(b) A major safety problem arises from the fact that the nuclear waste from the reactor contains some long-lived radioactive isotopes. Besides, appropriate cooling systems have to be designed to prevent accidents due to excessive heating (and melting) of the reactor core.

13.55 (a) $Q = [m(^2_1\text{H}) + m(^3_1\text{H}) - m(^4_2\text{He}) - m_n]c^2 = 17.59$ MeV.

(b) Repulsive potential energy of the two nuclei when they almost 'touch' each other

$$= \frac{q^2}{4\pi\epsilon_0 d}$$

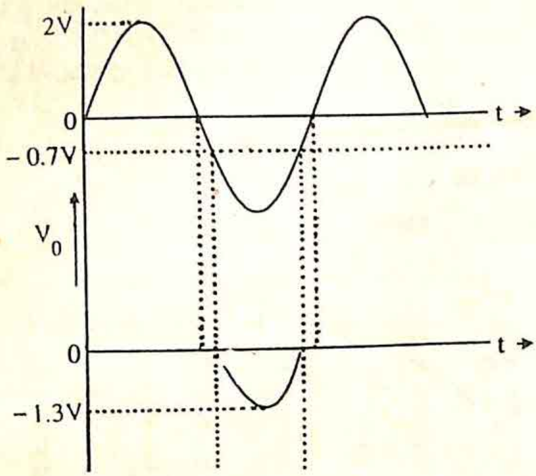
$$= \frac{9 \times 10^9 \times (1.6 \times 10^{-19})^2}{2 \times 1.5 \times 10^{-15}}$$

$$= 7.68 \times 10^{-14}\text{J.}$$

Classically, a kinetic energy at least equal to the above amount is required to overcome Coulomb repulsion. Using the formula K.E. = (3/2) kT, this corresponds to a temperature $T = 3.7 \times 10^9$K. Actually in practice fusion reactions are triggered at somewhat lower temperatures because of the quantum mechanical tunneling effect which need not concern us here.

Chapter 14

14.1

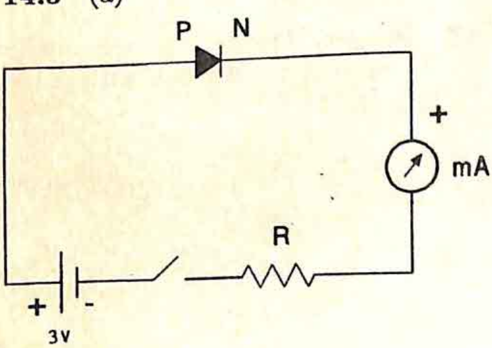


14.2 18.5 V; The Zener chosen should have a current rating of 17.5 mA and breakdown voltage should be 15V.

14.3 20 μA

14.4 200

14.5 (a)



(b) 0.046 W

14.6 Input frequency = 50 Hz.

Output ripple frequency after half wave rectification = 50Hz.

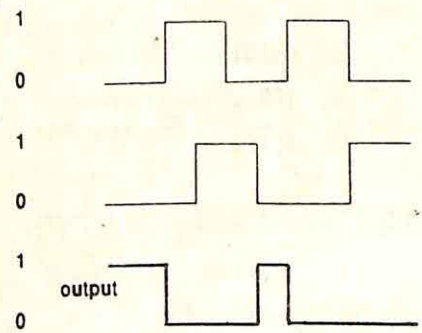
Output ripple frequency after full wave rectification = 100Hz.

14.7 Total Gain $G = G_1 G_2 = 20 \times 10 = 200$

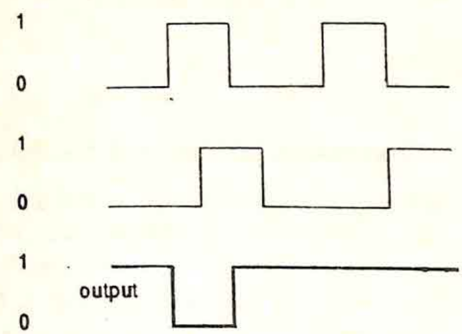
$$G = \frac{V_o}{V_i} = \frac{V_o}{50 \times 10^{-3}}$$

$$V_0 = 200 \times 50 \times 10^{-3} = 10 \text{ Volts.}$$

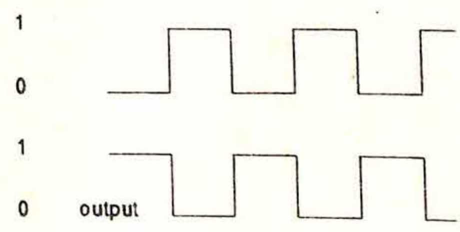
14.8
(a)



(b)



(c)



14.9 (a) 1101111

(b) 111111

(c) 1000000

14.10 (a) 001101

(b) 000010

(c) 010101

(d) 000000

14.11 (a) 0000001

(b) 1000000

(c) 010110

(d) 110111

14.12 (a) 001011

(b) 010101

(c) 0111110

14.13 $Y = A\bar{B} + \bar{A}B = A \oplus B (Ex - OR)$

TRUTH TABLE

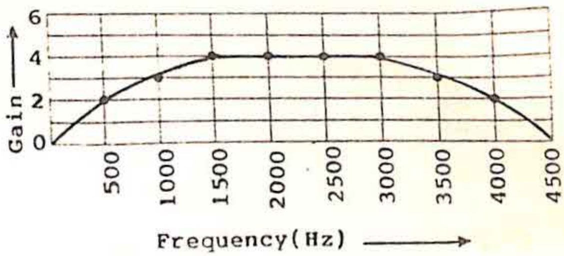
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Answers to Additional Exercises

14.14 The value of resistance required for the minimum voltage i.e. 10 V is 14.47 kΩ and for the maximum voltage, it is 6.64 kΩ. Therefore, by using a series combination of a variable resistor (7.83 kΩ) and a fixed resistor (6.64 kΩ), we can achieve the required variation at the output. The fixed resistor (6.64 kΩ) ensures that the voltage will not go above 20 V.

14.15 (a) V_0 at all the frequencies

(b)



(c) $f_1 = 1500$ Hz, $f_2 = 3000$ Hz.

14.16

A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

14.17 100 kHz. The JK flip is connected to work as a T flip flop with the input as 1.

Chapter 15

15.1 142, 873 km.

15.2 $\epsilon = 22.3^o$.

Distance between the earth and Mercury $= 1.384 \times 10^8$ km.

15.3 Smaller by a factor of 0.63.

15.5 1.563×10^{-11}

15.6 0.75.

15.7 158, 500.

15.8 Blue: 5912 K to 6438 K.

Red: 3762 K to 4673 K.

Answers to Additional Exercises

15.10 3,581 km.

15.11 2.31×10^{41} J.

15.12 About 374 times larger.

15.13 Density of the white dwarf $= 1.8 \times 10^{18}$ kg/km^3.

Density of the neutron star $= 4.78 \times 10^{26}$ kg/km^3 (which is nuclear density).

15.14 3.54×10^8 years.

15.15 $H = 15.12 (\text{km/s})/10^6$ ly;
$t_0 = 19.82 \times 10^9$ y.

BIBLIOGRAPHY

Text Books

For additional reading on the topics covered in this book, you may like to consult one or more of the following books. Some of these books however are more advanced and contain many more topics than this book.

(a) 'Ordinary Level Physics', by A. F. Abbott, 4th Edition, Arnold-Heinemann (1984).

(b) Advanced Level Physics by M. Nelkon and P. Parker, 5th Edition, Arnold-Heinemann (1982).

(c) Physics by T. Duncan, John Murray (1982).

(d) Physics, Parts 1 and 2, by D. Halliday and R. Resnick, 3rd Edition, John Wiley (1978).

(e) University Physics by H. D. Young, M. W. Zemansky and F. W. Sears, 6th Edition, Narosa Pub. House (1982).

(f) Problems in Elementary Physics, by B. Bukhovtsa, V. Krivchenkov, G. Myakishev and V. Shalnov, MIR Publishers, Moskow (1971).

(g) Lectures on Physics (3 volumes) by R. P. Feynman, Addison Wesley Pub. Co. (1965).

(h) Berkeley Physics Course (5 volumes) McGraw Hill (1965).

Vol. 1 - Mechanics: (Kittel, Knight & Ruderman)

Vol. 2 - Electricity and Magnetism (Purcell)

Vol. 3 - Waves and Oscillations (Crawford)

Vol. 4 - Quantum Physics (Wichmann)

Vol. 5 - Statistical Physics (Reif)

(i) Physics for the Technician by L. S. Zhdanov, MIR Publishers (1980).

(j) College Physics by R. L. Weber, K. V. Manning, M. W. White and G.A. Weygand, Tata McGraw Hill (1977).

(k) Physics: Foundations and Frontiers by G. Gamow and J. M. Cleveland, Tata McGraw Hill (1978).

(l) Physics for the Inquiring Mind by E. M. Rogers, Princeton University Press (1960).

(m) PSSC Physics Course: DC Heath & Co. (1965) Indian Edition, NCERT (1967).

General Books

For instructive and entertaining general reading on science, you may like to read some of the following books. Remember however, that many of these books are written at a level far beyond the level of the present book.

(a) Mr. Tompkins in paperback by G. Gamow, Cambridge University Press (1967).

(b) The Universe and Dr. Einstein by C. Barnett Time Inc. New York (1962).

(c) Thirty years that shook physics by G. Gamow Double Day, New York (1966).

(d) Surely You're joking, Mr. Feynman by R. P. Feynman, Bantam books (1986).

(e) One, Two, Three ··· Infinity by G. Gamow, Viking Inc. (1961).

(f) The Meaning of Relativity by A. Einstein, (Indian Edition) Oxford and IBH Publ Co. (1965).

(g) Atomic theory and the Description of Nature by Niels Bohr, Cambridge (1934).

(h) The Physical Principles of Quantum Theory by W. Heisenberg, University of Chicago Press (1930).

(i) The Physics – Astronomy Frontier – by F. Hoyle and J. V. Narlikar, W. H. Freeman (1980).

(j) The Flying Circus of Physics with Answers by J. Walker, John Wiley & Sons (1977).

(k) Physics for Everyone (series) by L. D. Landau and A. I. Kitaigorodski, MIR Publishers (1978).

Book 1: Physical Bodies

Book 2: Molecules

Book 3: Electrons

Book 4: Photons and Nuclei.

(l) Physics can be Fun by Y. Perelman, MIR Publishers (1986).

(m) Powers of Ten by Philip Morrison & Eames, W. H. Freeman (1985).

(n) Physics in your Kitchen lab. by I. K. Kikoin, MIR Publishers (1985).

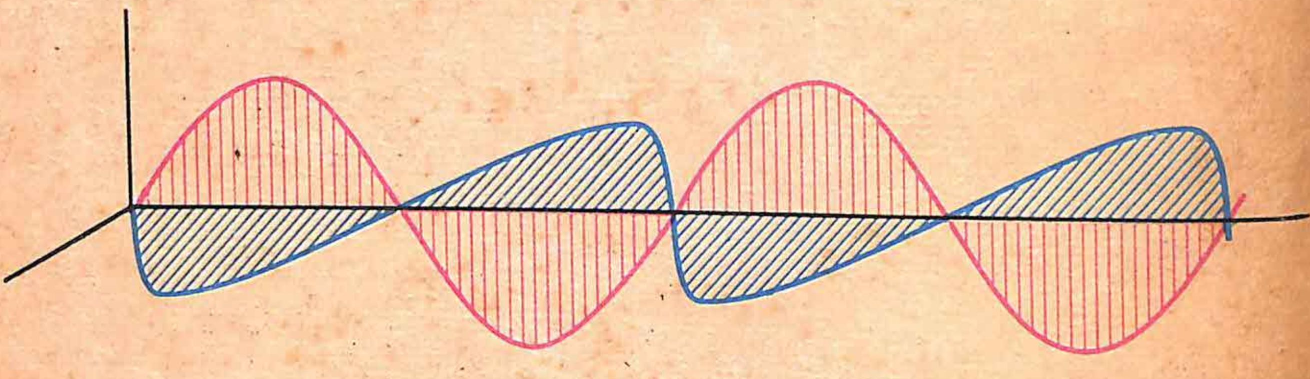
ERRATA FOR PHYSICS TEXTBOOK OF CLASS XII (PART I & II)

Page	*Column*	*Reference*	*Correction*
22	Left	Para 2, line 6	Replace σ by λ
25	Right	Fig. 1.17	Replace charge q on the right hand point by - q
44	Right	Example 2.1, line3	Replace 800 N/Cm^2 by 800 $N/C\ m^{1/2}$
77	Right	Fig. 2.35	The direction of external electric field be reversed as centre of positive charge will be displaced in the direction of the field
79	Left	Fig. 2.38	Delete 2. on the right side
108	Left	Line 14	Replace (zero degree Kelvin) by (zero kelvin)
117	Left	Line 6	Replace $Req = \varepsilon/I$ by $Req = \varepsilon/3I$
	Left	Fig. to Example 3.6	Replace the resistance 4Ω between points C and D by 2Ω
251	Right	Answer to Example 7.1, line 1-2	Replace 'Sonenoid' by 'Solenoid'
253	Left	Fig. 7.6	Interchange A^{++} and B^{--} and similarly D^{++} and C^{--}
	Left	Fig. 7.7	Interchange A^{++} and B^{--}. Replace D^{++} by C and C^{--} by D as there will be no charge accumulation on the ends C and D.
	Right	Fig. 7.8 (a)	Interchange ${}^{+}_{+}A^{+}_{+}$ and ${}^{-}_{-}B^{-}_{-}$
		Fig. 7.8 (b)	The direction of F needs to be reversed because charge on electron is negative
259	Right	Para 2, line 12	Replace 'kgs' by 'kg'
262	Right	Answer to Example 7.5, line 6	Replace $\omega = 314$ by $\omega = 314\ s^{-1}$

Page	*Column*	*Reference*	*Correction*
285	Right	R.H.S. of Eq. (8.34)	Replace I_m by Q_m
286	Right	Para 2, line 4	Replace ϕ by θ
304	-	Item 7, line 2	Replace $I = I_0 \sin(\omega t - \delta) = V_0 / \{R^2 + (\omega L - (1/\omega C))^2\}^{1/2}$ by $I = I_0 \sin(\omega t - \delta)$; $I_0 = V_0/\{R^2 + (\omega L - (1/\omega C))^2\}^{1/2}$
319	Right	Answer to Example 9.1, line 3 from bottom	Delete (comma) after the displacement current
355	-	Fig. 10.7	The electric vector in upward direction is E
356	Left	Fig. 10.8	Replace a⊥ by $a_\perp$ and a‖ by $a_\parallel$
370	Left	Answer to Example 10.6, line 1	Replace 5×10^{-7} by 5×10^{-7} m in the denominator of the equation
382	Left	Exercise 10.4, line 2	Read 'reflecing' as 'reflecting'
470	-	Line 7	Read $e = 1.602 \times 10^{-19}$ C in place of $e = 1.602 \times 19^{-19}$ C
471	-	Line 3	Replace (iv) by (d)
487	Left	Fig. 13.7	Replace Cot $\theta/2$ by cot $\theta/2$
560-562	-	Chapter 14	In figures, examples, exercises, read the resistance (given in k) as k Ω
687	Right	Line 3	Read Statistical Physics ( ) as Statistical Physics (Reif)

1270

राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING



Kaṇāda ★ Archimedes ★ Kepler ★ Galileo ★ Newton
Henry Cavendish ★ Charles Coulomb ★ Thomas Young
André-M Ampère ★ C F Gauss ★ Amedeo Avogadro
Michael Faraday ★ James P Joule ★ Lord Kelvin
J C Maxwell ★ Ludwig Boltzmann ★ J J Thomson
Max Planck ★ Marie Curie ★ Robert A Millikan
E Rutherford ★ Albert Einstein ★ Niels Bohr ★ de Broglie
Erwin Schrödinger ★ Werner Heisenberg ★ Paul Dirac
C V Raman ★ M N Saha ★ S N Bose ★ W Pauli
Enrico Fermi ★ Lev D Landau ★ S Chandrasekhar
Edwin Hubble ★ John Bardeen ★ Richard Feynman